

OM of: US-09-303-518d-125 to: SwissProt_40:* out_format: pfs
Date: Jun 30, 2002 8:31 AM

About: Results were produced by the GeneCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+np.model -DEV=ylh
-O=cnr2.1/USPRO.spool/US09303518/runat.28062002.142714.4362/app.query.fasta.1.23501
-DB=SwissProt_40 -QFMT=fastlan -SUFFIX=isp -GAPD=12.000
-GAPEX=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPD=4.500 -OGAPEXT=0.050 -XGAPD=10.000 -XGAPEXT=0.500
-FGAPD=6.000 -FGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS=human0.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09303518.@CGN1.1.440 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-125
Query length: 1344
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000

score list:

Sequence	Strid	Orig	Zscore	Escore	Len	Documentation
SwissProt_40:NORA_NEIMA	2289.00	3051.97	6.0e-163	447	09K0M3	neisseria meningitidis
SwissProt_40:NORA_NEIMA	2251.00	3051.15	4.0e-160	447	09JVP8	neisseria meningitidis
SwissProt_40:NORA_PASMU	1676.50	2232.82	2.5e-117	446	09C4B1	pasteurella multocida
SwissProt_40:NORA_HEIN	1629.00	2169.27	8.8e-114	447	09A395	haemophilus influenzae
SwissProt_40:NORA_VIBHA	1471.50	1958.64	4.7e-102	446	09F102	vibrio haeveryi na(++)
SwissProt_40:NORA_VIBAL	1468.50	1954.64	7.9e-102	446	09F586	v na(+)-translocating
SwissProt_40:NORA_VIBCH	1455.50	1937.24	7.4e-101	445	09KPS1	vibrio cholerae na(++)
SwissProt_40:NORA_PSEAE	1353.00	1800.18	3.2e-93	445	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_CHLNU	471.00	620.14	1.6e-27	465	09JVP8	chlamydia muridarum
SwissProt_40:NORA_CHLNU	464.50	611.41	5.0e-27	465	09JVP8	chlamydia pneumoniae
SwissProt_40:NORA_VIBCH	158.00	196.70	0.0004	774	09A488	chlamydia trachomatis
SwissProt_40:NORA_VIBCH	150.00	189.79	0.0014	519	052716	rhodospirillum rubrum
SwissProt_40:NORA_VIBCH	138.50	170.09	0.0109	819	071397	haemophilus influenzae
SwissProt_40:NORA_VIBCH	138.00	161.96	0.0140	1802	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_VIBCH	129.50	154.44	0.0552	1199	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	129.00	157.92	0.0547	774	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	126.00	158.57	0.0824	473	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	125.00	160.69	0.0905	328	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	124.00	147.56	0.1402	1140	09A488	chlamydia trachomatis
SwissProt_40:NORA_PSEAE	115.00	139.62	0.5985	740	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	113.00	136.65	0.7910	819	09JVP8	haemophilus influenzae
SwissProt_40:NORA_PSEAE	112.50	140.65	0.8324	466	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	112.00	140.59	0.8945	437	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	111.50	139.90	0.9751	438	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	110.50	133.85	1.61	2073	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	107.00	123.11	2.69	1367	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	107.00	123.11	2.69	1367	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	105.50	125.98	3.12	817	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	105.50	115.76	3.93	2404	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	105.00	130.81	3.00	457	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	104.50	129.65	3.31	481	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	104.50	124.59	3.71	821	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	104.50	118.22	4.28	1609	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	103.50	128.31	3.92	481	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	102.50	127.28	4.63	466	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	102.50	121.76	5.24	835	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	102.50	118.10	5.69	1229	09JVP8	pseudomonas aeruginosa
SwissProt_40:NORA_PSEAE	102.00	127.61	4.93	419	09JVP8	pseudomonas aeruginosa

SwissProt_40:OD2_ZYMO	102.00	127.15	4.98	440	066119 zymomonas mobilis
SwissProt_40:OD2_ZYMO	102.00	119.35	5.94	1003	061872 zymomonas mobilis
SwissProt_40:OD2_ZYMO	102.00	112.40	6.95	2090	051611 mesocricetus aure
SwissProt_40:OD2_ZYMO	101.50	122.33	5.96	682	050478 gallus gallus (chi
SwissProt_40:OD2_ZYMO	101.00	126.58	5.81	406	077444 escherichia coli
SwissProt_40:OD2_ZYMO	101.00	121.46	6.52	697	071886 mycobacterium tube
SwissProt_40:OD2_ZYMO	101.00	115.52	7.46	1306	072334 saccharomyces cere
SwissProt_40:OD2_ZYMO	100.50	123.79	6.64	508	010944 bacillus subtilis
SwissProt_40:OD2_ZYMO	100.50	120.42	7.03	725	023323 saccharomyces cere
SwissProt_40:OD2_ZYMO	100.00	120.90	7.61	2139	001815 mus musculus (mou
SwissProt_40:OD2_ZYMO	100.00	116.81	8.34	642	022599 saccharomyces cere
SwissProt_40:OD2_ZYMO	99.50	120.41	8.26	630	060550 mesocricetus aure
SwissProt_40:OD2_ZYMO	99.50	118.19	8.68	797	028968 equine herpesvirus
SwissProt_40:OD2_ZYMO	99.00	106.79	12.05	2476	028983 sus scrofa (pig) -
SwissProt_40:OD2_ZYMO	98.50	123.66	8.84	388	052424 vigna unguiculata
SwissProt_40:OD2_ZYMO	98.00	123.29	9.56	376	060077 saccharomyces cere
SwissProt_40:OD2_ZYMO	98.00	122.45	9.75	411	067049 schizosaccharomyce
SwissProt_40:OD2_ZYMO	97.50	116.81	11.88	695	049059 sus scrofa (pig) -
SwissProt_40:OD2_ZYMO	97.00	120.57	11.71	435	039815 bacillus subtilis
SwissProt_40:OD2_ZYMO	96.00	112.74	16.10	864	091912 homo sapiens (huma
SwissProt_40:OD2_ZYMO	95.50	120.30	14.56	362	091912 homo sapiens (huma
SwissProt_40:OD2_ZYMO	95.50	117.06	15.67	510	013634 simian immunodefici
SwissProt_40:OD2_ZYMO	95.50	114.13	16.74	695	032212 macaca fasciculari
SwissProt_40:OD2_ZYMO	95.50	113.55	16.97	739	015143 gallus gallus (chi
SwissProt_40:OD2_ZYMO	95.50	109.76	18.48	1103	014859 homo sapiens (huma
SwissProt_40:OD2_ZYMO	95.50	106.31	19.98	1589	054210 diuretic acidog
SwissProt_40:OD2_ZYMO	95.00	115.94	17.25	535	039769 drosophila melanog
SwissProt_40:OD2_ZYMO	95.00	115.36	17.48	569	006219 mycobacterium tube
SwissProt_40:OD2_ZYMO	95.00	103.29	22.96	2035	008719 helicobacter felis
SwissProt_40:OD2_ZYMO	94.50	102.69	23.27	2169	022002 rattus norvegicus
SwissProt_40:OD2_ZYMO	94.50	114.84	18.98	488	024310 pisum sativum (garde
SwissProt_40:OD2_ZYMO	94.50	114.44	20.53	802	052271 rickettsia prowaz
SwissProt_40:OD2_ZYMO	94.50	106.34	23.00	1374	037218 lycopersicon esculen
SwissProt_40:OD2_ZYMO	94.00	119.32	18.41	325	032934 bluetongue virus (s
SwissProt_40:OD2_ZYMO	94.00	115.28	20.17	498	026530 nicotiana rustica
SwissProt_40:OD2_ZYMO	94.00	109.84	22.81	885	000826 nicotiana tabacum
SwissProt_40:OD2_ZYMO	93.50	115.18	21.69	469	049574 odontella sinensis
SwissProt_40:OD2_ZYMO	93.50	113.92	22.32	536	040996 schizosaccharomyce
SwissProt_40:OD2_ZYMO	93.50	113.71	22.43	548	051772 pseudomonas fluores
SwissProt_40:OD2_ZYMO	93.50	106.60	26.33	1161	047179 saccharomyces cere
SwissProt_40:OD2_ZYMO	93.50	105.45	27.03	1311	099495 drosophila melanog
SwissProt_40:OD2_ZYMO	93.00	101.04	23.86	2090	035658 homo sapiens (huma
SwissProt_40:OD2_ZYMO	93.00	114.27	23.76	481	059821 staphylococcus aure
SwissProt_40:OD2_ZYMO	93.00	114.10	23.86	490	091447 buchnera aphidicola
SwissProt_40:OD2_ZYMO	93.00	113.94	23.94	498	026529 nicotiana glauca
SwissProt_40:OD2_ZYMO	93.00	113.44	24.21	525	017955 rattus norvegicus (r
SwissProt_40:OD2_ZYMO	93.00	110.72	25.75	700	045537 escherichia coli
SwissProt_40:OD2_ZYMO	93.00	110.04	26.15	752	054351 drosophila melanog
SwissProt_40:OD2_ZYMO	93.00	109.82	26.28	770	078994 mycobacterium tube
SwissProt_40:OD2_ZYMO	93.00	108.54	27.05	881	036016 saccharomyces cere
SwissProt_40:OD2_ZYMO	93.00	107.41	27.75	993	040694 mus musculus (mous

seq_name: SwissProt_40:NORA_NEIMA
seq_documentation_block:
ID NORA_NEIMA STANDARD: 447 AA.
AC 09K0M3: 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
subunit A) (NQR-1 subunit A).
GN NORA OR NM00569.

OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,
 RA Nelson W.C., Gwin M.U., Deboy R., Peterson J.D., Hickey E.K.,
 RA Hait D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Checko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 RA Cotton M.D., Ufferback T.R., Khouri H., Olin H., Vamathevan J.,
 RA Gill J., Scarlato V., Masiagni V., Pizsa M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
 CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
 CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORE
 CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
 CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
 CC UBIQUINOL + NA(+)(OUT).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NOB, NOC, NOD, NOE
 CC AND NOF (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE002412; AAF40997.1; -;
 DR TIGR; NMB0569; -;
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 KW Complete proteome.
 SQ SEQUENCE 447 AA; 48635 MW; BE17A7A8439BE477 CRC64;

alignment_scores:
 Quality: 2289.00 Length: 447
 Ratio: 5.121 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-303-518D-125 x NORA_NEIMB ..
 Align seg 1/1 to: NORA_NEIMB from: 1 to: 447

```

1 ATGATTAATCAAAAAGTCTAAAGTCCCATCGCGGCGACGCGGA 50
|||||
1 MetIleTylsIleTylsGlyLeuAsnLeuProIleAlaGlyArgProG1 17
51 GGAAGCCGTTTACGAGCGCGCGCATTCAGCAAGTCCGCTTGGCG 100
|||||
17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34
101 AAGAATATGCGCGTATGCGCCCTCGATGAATCAAGCAAGCGCGATGCC 150
|||||
34 IuGluTyrAlaGlyMetArgProSerMetTylsValTylsGluGlyAspAla 50
151 GTCAAAAAAGGCCAAGTGTCTTTGAAGCAAAAAGAAATCCGGCGGTGT 200
|||||
51 ValTylsTylsGlyGlnValLeuPheGluAspTylsTylsAsnProGlyValAla 67
201 GTTTACTGGCGCGCGCTCAGCAAAATGCGCGCATTCACCGTGGCGAA 250
|||||
67 lPheThrAlaProAlaSerGlyTylsIleAlaIleHisArgGlyGluL 84

```

```

251 AGCGCTACTTCAGTCACGTCGATTTGCGGTGAAGCAACGCAAAATC 300
|||||
84 YsArgValLeuGlnSerValValIleAlaValAlaGluGlyAsnAspIle 100
301 GAGTTTGAACGCTACGACACCTGAGCGGTGCGCAAACTTAAGCGCGA 350
|||||
101 GluPheGluArgTyrAlaProGluAlaIleAlaAsnLeuSerGlyGlu 117
351 AGTGGCGCGCAACCTGATTCATCCGCTTGGCGCTGCGCGCGCACCC 400
|||||
117 uValAlaGluAsnLeuIleGlnSerGlyLeuTyrPheAlaLeuArgTyr 134
401 GTCGCTTCGCAAAATTCCTGCGCGTATGCGCGCGCGCGCATTCCT 450
|||||
134 rGProPheSerTylsIleProAlaValAlaAspAlaGluProPheAlaIlePhe 150
451 GTCAATGCGATGAGCAACCAATCCGCTGCGTGGCGCAACCTACGTCAT 500
|||||
151 ValAsnAlaMetAspThrAsnProLeuAlaIlaAspProThrValIle 167
501 CAAAGAACGCCCGGAGAGATTTCGAACGCGCGCTGTTGATGAGCGGT 550
|||||
167 eTylsGluAlaAlaGluAspPheTylsArgGlyLeuLeuValLeuSerArgL 184
551 TGACCGAACGCAAAATCCATGTTTGAAGCAGCTGCGCGCAGACGTGCC 600
|||||
184 eThrGluArgTylsIleHisValCysTylsAlaAlaGlyAlaAspAlaPro 200
601 TCTGAATAATGCTGCCAATCGAACACATGATTTGGCGCGCGCATTC 650
|||||
201 SerGluAsnAlaIleAsnIleGluThrHisGluPheGlyLProHisr 217
651 TGCCGGTTGAGTGCAGCGCACATTCATTCATTCGACGCGGTGCGCGCA 700
|||||
217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValAla 234
701 ATAAACCGGTGGACCATCAATTATCAAGATGTAATTCATTGGCGCT 750
|||||
234 snTylsThrValTyrPheIleAsnTylsGluAspValIleThrIleGlyArg 250
751 TTGTTTGAACAGCGCGTCTGAACACGAGCGCGGTGATTCGCTAGGTG 800
|||||
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGly 267
801 TTCTCAAGTCACAAACCGCGCTCTGCGTACCGCTTTGGTGGCGAAG 850
|||||
267 YSerGlnValAsnTylsProArgLeuLeuArgThrValLeuGlyAlaTylsV 284
851 TATCGCAATTTACTGCGCGCGAATTGGTTGACACAGCAACCGCGTGAT 900
|||||
284 alSerGlnIleThrAlaGlyGluLeuValAspThrAsnAsnArgValIle 300
901 TCCGCTTCGCTATGAACGCGCGATTCACCAAGCGCGCGCAGATATT 950
|||||
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyls 317
951 GCGACGCTACCAATCAGATTTCGTTATCGAAGAGCGCGCAGCAAG 1000
|||||
317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerTylsG 334
1001 AGCTGTTGCGGTGGTGGCGCGCGCAGCGCAAAATACCTCATCGCGCT 1050
|||||
334 lLeuPheGlyTyrValAlaProGlnProAspTylsTylsSerIleThrArg 350
1051 ACAACCGTGGCGATTCCTGAAAAACAACTTCGATCAAGTTCACACAGC 1100
|||||
351 ThrThrLeuGlyHisPheLeuTylsAsnTylsLeuPheTylsPheAsnThrAl 367
1101 CGTCAACGCGCGCGCGCGCATGTGTGCGGATTTGATTCAGACGCGG 1150
|||||
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTylsGluArgV 384

```

```

1151 TGATGCTCCCTGGATATCTGCCACCTGCTTTGGCGGATTAATCTGTC 1200
|||||
384 almetProleuspileleuProthleuLeuAArgAspleuileval 400
1201 GGCGATACCGACAGCGGAGCATTTGGTTGGAATTGGACGAGA 1250
|||||
401 GYAAPTThraspSerlaInlaLeuGlyCysLeuGluLeuAspGluGI 417
1251 AGACCTGCTTTGTCAGCTTGTGCTGCGCGGCAAAATAGCAATAGCGGC 1300
|||||
417 uaspleuAlaLeuGlyCysSerPheValCysProGlyLysTyrGluTyrGI 434
1301 CGCTGTGCGCAAGTGTCTGGAACCATTTAGAGAGAGGC 1341
|||||
434 rleuLeuAArgLysValLeuGluThrIleGluLysGluGI 447

```

seq_name: SwissProt_40: NORA_NEIMA

seq_documentation_block:

```

ID NORA_NEIMA STANDARD; PRT; 447 AA.
AC O9JVP8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
subunit A) (NQR-1 subunit A).
GN NORA OR NMA0752.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
ON NCBI_TaxID:55699;
RX STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE-20222556; PubMed-10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jørgels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQR
ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
UBIQUINOL + NA(+) (OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NORB, NQRC, NQRD, NQRE
AND NQRF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AL162754; CAB84035.1;
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
CC Complete proteome.
CC KW
SQ SEQUENCE 447 AA; 48673 MW; B808CAA82C901A1D CRC64;

```

alignment_scores:

Quality: 2251.00 Length: 447
 Ratio: 5.036 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 97.763

alignment_block:
 US-09-303-518d-125 x NORA_NEIMA

Align seq 1/1 to: NORA_NEIMA from: 1 to: 447

```

1 ATGATTAAATCAAAAAAGGTCTAAACCTGCCATCGCGGCGAGACGGA 50
|||||
1 MetIleLysIleLysIleLysGlyLeuAsnLeuProIleAlaGlyArgProI 17
51 GCAGCCGTTTACGACGCGCGGCATTTACCGAAGTCCGCTGCTGGCG 100
|||||
17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGI 34
101 AAGAAATGCGCGGTATGCGCCCTCGATGAATCAAGAGAGCGATGCG 150
|||||
34 LuGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50
151 GTCAAAAAGCCCAAGTGTCTTGAAGACAAAAAAGATCCGGCGTGT 200
|||||
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 GTTTACTGCGCGGCTTCAAGCAAAATCGCGCGATTCACCGTGCGAAA 250
|||||
67 LpheThrAlaProValSerGlyLysIleAlaIleHisArgGlyGI 84
251 AGCGCGTACTTCAGTCACTGCTGATTCGCTTGAAGCAACGACGAATC 300
|||||
84 YsaArgValLeuGlnSerValIleValAlaGluValGluLysAsnArgIle 100
301 GACTTTGAACGCTACGACGACCTGAAGCGCTGGCAAACTTGAAGCGGAGAGA 350
|||||
101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGI 117
351 AGTGGCGCGCAACCTGATCCAAATCCGCTTGTGACTGCGTCCGACCC 400
|||||
117 uValArgArgAsnLeuIleGlnSerGlyLeuTyrPheAlaLeuArgThr 134
401 GTCCGTTACCAAAATTCCTGCGCGTGAAGCCGACGCGTTCGCCACTTC 450
|||||
134 rGProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150
451 GTCAAATGCATGAGACACAATCCGTGCGCGCGACCTGATCAATAT 500
|||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValIle 167
501 CAAAGAACCGCGGAGGATTTCAACGCGCGCTGTTGATTTAGCGGTT 550
|||||
167 eLysGluAlaAlaGluAspPheArgGlyLeuLeuValLeuSerArgL 184
551 TGACCGAAGCAAAATCCATGTTTGAAGCAACCTGGCGCAGACGTGGC 600
|||||
184 eThrGluArgLysIleHisValCysLysAlaIleGlyAlaAspValPro 200
601 TCTGAAATGCTGCCAATCGAACAACATGATTTGGCGGCGCGGATCC 650
|||||
201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyProHisr 217
651 TGCGCGTTGATGAGCAGCAGCATTCATTCATGAGCGCGTGGCGGA 700
|||||
217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAla 234
701 ATGAAACCGTGTGACATCAATTAACAAGTATTAACATTTGGCGCT 750
|||||
234 snLysThrValIleThrIleAsnTyrGlnAspAlaIleAlaIleGlyArg 250
751 TTCTTGAACAGCGCGTCTGAACACGAGCGCGGATTTGCCCTAGGTGG 800
|||||
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGI 267
801 TTCTCAAGTCAACAACCGCGCTTGGGTACCGCTTTGGGTGGGAAG 850
|||||
267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

```

```

851 TATCGCAATTAAGTGGGGGCGGATGTTGACACAGACACCGCGGATT 900
284 aIserrhnllethrrAlaaglyleuValaspaIaspsnAtgValle 300
901 TCCGGTTCGGTATTGAACGGCGGATTAACAAGCGCGCGCATATTAT 950
301 SerGlySerValleuasnGlyAlaIlethnInGlyAlaHisapryle 317
951 GGGACGCTACACATCAATTCCTTCCTATCCAGAAAGCGCGCAAG 1000
317 uGlyAgtYrHisasnGlnIleSerValIleGluGlyAgtSerlyG 334
1001 AGCTGTTCGGCTGGTGGCGCCGACCGGACAAATATCCATCAGCGT 1050
334 lueuRhegIyTTPValAlaIaProGlnProAspIySTySerIlethrrAl 350
1051 ACAACCTCGGCGCATTCCTGAAAAACAACCTTCAGTCAACACAGC 1100
351 ThrThleuGlyHisPheleuLysAsnLysleuPheLysPheThrrAl 367
1101 CGTCAACGGGGGCGACCGCCCATGGTGGCGGATGTTACTTACAGCGG 1150
367 aValasnGlyGlyAspArgAlaMetValProIleGlyThrYrGlyAgt 384
1151 TGATGCCCTTGATATCTGCTGCCACCCGCTTTGGCGGATTAATGTC 1200
384 alMetProleuAspIleleuProThrleuLeuArgAspleuIleVal 400
1201 GCGGATACGACAGCGCGCGCATGGTGGTGGTGAATTGGACGAGA 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGlnleuAspIle 417
1251 AGACCTCGCTTGTGAGCTTGTGCTGCCCGGCAATAGATACGGCG 1300
417 uAspleuAlaLeuCysSerPheValCysProGlyIySTyGlyYrGly 434
1301 CGCTGTTCGCAAAAGTCTGGAACCATTTGAGAAGAGAGGC 1341
434 roleuLeuArglyValleuGlnThrIleGlnLysGlnly 447

```

seq_name: SwissProt_40: NORA_PASMU

```

seq_documentation_block:
ID NORA_PASMU STANDARD: PRT: 446 AA.
AC 09CBL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE Na(+)-translocating NOR subunit A (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR PM1328;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70."
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -I- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORA
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -I- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE
CC AND NORF (BY SIMILARITY).

```

-I- SIMILARITY: BELONGS TO THE NORA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE006171; AK03412.1; -
 KM Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 KW Complete proteome.
 SQ SEQUENCE 446 AA; 48168 MW; 5C5287CB77C4932D CRC64;

alignment_scores:
 Quality: 1676.50 Length: 447
 Ratio: 4.150 Gaps: 1
 Percent Similarity: 90.380 Percent Identity: 71.588

alignment_block:
 US-09-303-518d-125 x NORA_PASMU ..
 Align seg 1/1 to: NORA_PASMU from: 1 to: 446

```

1 ATGATTAATAATCAAAAAAGCTTAACCTGGCCATGCGGCGAGACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 MetIlethrrIleLysGlyLeuAsnleuProIleSerGlySerProG 17
51 GCAAGCCGCTTTACGAGCGCGCGCCGATTAACGAGTGGCGTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 uGlnValIleArgAspGlyAsnAlaIlethrrGlnValAlaIleuG 34
101 AAGATATAGCGCGGTATGGCGCCCTCGATGAAGTCAAGAGAGCGAT 150
||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 uGlnIyTValGlyMetArgProSerMetLysValArgGlnGlyAsp 50
151 GTCAAAAAGGCCAAGTCTGTTGAAGACAAAAGATCGGCGCGTGGT 200
||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 ValIyLysGlyGlnValleuPheGlnAspLysAsnProGlyValIa 67
201 GTTACTGCGCGCGCTTCAGGCAAAATCGCGCGGATTCACCGTGGCAA 250
||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 lPheThrrAlaProAlaSerGlyThrValThrAlaIleHisArgGly 84
251 AGCGGTACTCTACAGTGTGATGCTGATGCGCTGGCAAGCAACGAAATC 300
||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ySarGValleuGlnSerValIleLysIleGlnGlyAsnGlnIle 100
301 GAGTTTGAAACGCTACGACCTGAGAGCGCTGGCAAACTTAAGCGCGAGA 350
||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ThrPheGlnLysIyTThrThrrGlnleuAsnIleuThrSerGln 117
351 AGTGGCGCGCACTGATCCATCCGGTTGTGTGACTGCGTGGCACCC 400
||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 nValArgGlnAsnleuGlnAlaSerGlyLeuThrPThrAlaLeuArg 134
401 GTCCGTTAGCAAAATTCCTCGCTCGATGCGGAGCGCGTTCGCATCTTC 450
||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 rGpProPheSerLysValProAlaValAspAlaThrProValSer 150
451 GTCAATGAGTATGACACCAATCCGCTGGTGGCGGACCTGACGATTAAT 500
||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValAsnAlaMetAspThrAsnProleuCysAlaAspProGlnValI 167
501 CAAAGACCGCGAGGATTTCAACCGCGCGCTGTTGATGAAGCGGTT 550
||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 lGlnIleSerAlaGlnAlaPheGlnValGlyLeuThrValleuSer 184
551 TGACCGAAGCGAAATTCATGTTGTGAAGGACGTGGCGGACAGACGTC 600
||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 euHisGlnGlyLysValIyTleuCysAlaIleAsnAlaSerIlePro 200

```


alignment_scores:

Quality: 1629.00 Length: 448
 Ratio: 4.093 Gaps: 2
 Percent Similarity: 88.839 Percent Identity: 68.973

alignment_block:

US-09-303-518D-125 x NORA_HAEIN ..

Align seg 1/1 to: NORA_HAEIN from: 1 to: 447

```

1 ATGATTAAATCAAAAGATCTAAACCTGACCATGCGGCGAGACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyLysProI 17
51 GCAAGCCGTTTACGAGCGCGCCGCAATTCAGAGTCGCGTGTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 aglnValIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGly 34
101 AAGAATATGCCGATATGCGCCGCTGATGAAAGTCAAGAGCGCATGCC 150
||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 lueGlyValGlyMetArgProSerMetLysValArgGlyLysArgVal 50
151 GTCAAAAAGGCCCAAGTCGTTTGAAGACAAAAGAAATCCGGGGTGGT 200
||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 ValLysLysGlyGlnValIleuPheGlnAspLysLysAsnProGlyVal 67
201 GTTTACTGCGCGGCTTTCAGGCAAAATCGCGCATTCACCGTGGCGAA 250
||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 epherThrAlaProIleSerGlyThrIleThrAlaIleAsnArgGlyLul 84
251 AGCGGCTACTGATGTCAGTCGATGATGCGGTGAAGGCAAGCAACATC 300
||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ysaArgValIleuGlnSerValValIleAsnValGlnGlyAspGlyLys 100
301 GAGTTTGAACCTACGACCACTGAGCGCTGCAAACTTAAGCGCGAAGA 350
||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ThrPheAlaLysTyrSerThrGlnIleuAsnThrLeuSerSerGlnG 117
351 AATGCGCCGCAACCTGATCCAAATCGGTTTGTGACTGCGCTGCGACCC 400
||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 nValLysGlnAsnLeuIleGlnSerGlyLeuThrPheAlaLeuArgThr 134
401 GTCCGTTGACGAAATTCCTGCGCTGATCCGAGCCGCTGGCCATCTTC 450
||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 rgiProPheSerLysValProSerIleGlnSerGlnAlaSerSerIlePhe 150
451 GTCAATGCGATGACACCAATCGCTGGTGGCGGACCGCATATAT 500
||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValVal 167
501 CAAGAAGCGCCGAGGATTTCAACGCGGCTGTTGATTTGAGCCGCT 550
||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 ulysGlyTyrSerGlnAspPheThrAsnGlyLeuThrAlaLeuSerArg 184
551 TACAGCAAGCGCAA...ATCCATGTTTGAAGGACGCTGCGCGAGAGCTG 597
||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 eurPheProSerLysProLeuHisLeuGlyLysAlaGlyAspSerAsnIle 200
598 CCGTGTGAATAATGCTCCCAACATCGAACAACATGATGATGCGGCGCC 647
||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ProThrAlaAspLeuGlnAsnLeuGlnIleHisAspPheThrGlyVal 217
648 TCCGTCGCGGTTTGAAGTGCAGCAATTCATTTATCGAGCGGTCGCG 697
||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 sProAlaGlyLeuValGlyThrHisIleHisIlePheLeuAspProVal 234
698 CGAATTAACCGCTGTGACATCAATTTATCAAGATGATTAATCAATG 747
||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 leGlnLysThrValThrHisIleHisnLysnLysArgValIleAlaVal 250
748 CGTTTGTGCAAGCGCGCTGTGCAACGAGCGCGGCGGATGGCCCTAG 797
||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 LysLeuPheThrThrGlyIleuLeuTyrSerGlnArgValIleSerLeu 267

```

```

798 TGGTTCTCAAGTCAACAACCGCGCTCTTGCGTACCGCTTTGGGTGCGA 847
||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 aglyProGlnValLysGlnProArgLeuValArgThrIleGlyAla 284
848 AAGTATCGCAAAATTCAGCGCGCAATTCGTTGACACAGCAACCGCGTG 897
||||| ||||||| ||||||| ||||||| ||||||| |||||||
284 snLeuSerGlnLeuThrGlnAsnGlnLeuSerAlaGlyLysAsnArgVal 300
898 ATTTCCGCTTCGATTTGAACGCGCGCATTCACAGCGCGCGCATTTA 947
||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 lIleSerGlySerValLeuGlyGlnIleAlaLysAspSerHisAspTy 317
948 TTTGAGACGCTTACCAATCAGATTCCTTATCGAAGAGCGCGCGACA 997
||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 rLeuGlyArgThrAlaLeuGlnValSerValIleAlaGlnGlyAsnGln 334
998 AAGAGCTGTTCGCGCTGCTGCGCGCGAGCGGCAAAATTCATCAGCAG 1047
||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 ysgLysPhePheGlyTyrPheMetProGlnAlaAsnLysTyrSerVal 350
1048 CGTCAACCGCTCGCGCATTCCTCGAATAACACTCTCAAGTTCAAC 1097
||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 ArgThrValLeuGlnLysPheSerLys...LysLeuPheAsnPheThr 366
1098 AGCGCTCAAGCGCGCGCGCGCGCGCATGTCGCGATTTGATTAAGAC 1147
||||| ||||||| ||||||| ||||||| ||||||| |||||||
366 rSerGlnAsnGlyLysGlnArgAlaMetValProIleGlySerTyrGln 383
1148 GCGTGATGCCCTTGATATCTGCGCACCTGCTTTGGCGCATTAATC 1197
||||| ||||||| ||||||| ||||||| ||||||| |||||||
383 rValMetProLeuAspIleLeuProThrLeuLeuArgAspLeuIle 399
1198 GTGCGGATACCGACAGCGCGAGCATTTGGTGTGGAATTTGACGA 1247
||||| ||||||| ||||||| ||||||| ||||||| |||||||
400 ValGlyAspThrAspGlyAlaGlnGlnLeuGlyLysLeuIleuAspG 416
1248 AGAAGACCTCGCTTTGTCAGCTTCGTCGCGCGCGCAATACGATACG 1297
||||| ||||||| ||||||| ||||||| ||||||| |||||||
416 uGlnAspLeuAlaLeuGlySerPheValLysProGlyLysTyrGlyTyrG 433
1298 GCCCGCTGTTGCGCAAGTGTGGAACCAATGAGAGGAGCGC 1341
||||| ||||||| ||||||| ||||||| ||||||| |||||||
433 LysrIleLeuArgGlnValLeuAspLysIleGlnLysGlnGly 447

```

seq_name: swissprot_40:NORA_VIBHA

seq_documentation_block:

ID	NORA_VIBHA	STANDARD:	PRY:	446 AA.
AC	O9RFW1;			
DT	16-OCT-2001 (Rel. 40, created)			
DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	16-OCT-2001 (Rel. 40, last annotation update)			
DE	Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)			
DE	(Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex subunit A) (NOR-1 subunit A).			
DE	NORA.			
OS	Vibrio harveyi.			
OC	Bacteria; Proteobacteria; gamma subdivision: Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=69;			
RN	111			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BB120;			
RA	MEDLINE-20056044; PubMed-10587447;			
RA	Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,			
RA	Genis R.B., Bogachev A.V., Barquera B.;			
RT	"Sequencing and preliminary characterization of the Na ⁺ -translocating			
RT	NADH:ubiquinone oxidoreductase from Vibrio harveyi."			
RL	Biochemistry 38:16246-16252(1999).			
CC	-I- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO			
CC	UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT			
CC	OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE			
CC	ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF			

CC UBISEMIQUINONE TO UBIQUINOL.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
 CC UBIQUINOL + NA(+) (OUT).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NQRA, NQRB, NQRC, NQRD, NQRE
 CC AND NQRF (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NQRA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF165980; AF15411.1; -
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
 SQ SEQUENCE 446 AA; 48365 MW; 9B05B38BBD7A97C CRC64;

alignment_scores:
 Quality: 1471.50 Length: 447
 Ratio: 3.812 Gaps: 1
 Percent Similarity: 86.353 Percent Identity: 63.087

alignment_block:
 US-09-303-518D-125 x NQRA_VIBHA ..

Align seg 1/1 to: NQRA_VIBHA from: 1 to: 446

```

1 ATGATTAATAATCAAAAAAGGCTTAACCTGACCCATCGCGGCGAGACCGGA 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIethrIleIysIysGlyLeuAlaIeuProIleAlaIleAlaProIse 17
51 GCAACCGCTTACGACGCGCGCCGCTTACCGAAGTCCGCTTGGTGGCG 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 rGlIValIleAsnAspGlyIysThrIleIysIleValAlaIleuIeuIlg 34
101 AAGATATGCGGATGCGCGCTGATGATGATGATGATGATGATGATGATGAT 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 IuGIuTyValIglYmeTarGProthIethIAsValArgValAlaIAspIu 50
151 GTCAAAAAAGGCGCAAGTCTGTTTGAAGACAAAAAGAAATCCGGCGTGT 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValIysIysAlaIglValIleuPheIgluAspIysAsnProIglValIly 67
201 GTTTACTGCGCGCGCTTACGAGCAAAATCGCGCGATTCACCGTGGCGAA 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 sPheThrIaProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 84
251 AGCGGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysaIrgValIeuIglIserIaValIleIgluValAlaIaIglIgluIgluI 100
301 GAGTTTGAACGCTTACGACCTGTAAGCGCTGGCAAACTTAAGCGCGAAGA 350
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ThrPheAspIysPheIgluAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 117
351 AGTGGCGCGCAACCTGATCAATCGGTTTGGTGGTGGTGGTGGTGGTGGT 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 IleIysThrIgluIeuValIgluIserGlyIeuTrpThrAlaIeuAlaI 134
401 GTCCGCTGCAAAATCTGCGCTGCGTGAAGCGGACCGTGGCGCATCTTC 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rGpProPheSerIysValIProIaIaIaIaIaIaIaIaIaIaIaIaIa 150
451 GTCAATGCGATGACCAATCGCTGGCTGGCGACCGCTAGCGCATTTAT 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValThrAlaIeIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 167
501 CAAAGAGCGCGCGGAGATTTCAACGCGGCGCTGTGGTATTTGAGCGCT 550
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eThrIgluIgluIgluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 184

```

```

551 TGACGAGCAACAAATCATGTTTGAAGCAGCTGGCGAGCGATGCGG 600
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 eutHrIgluIgluIysValItyrValCysIys...SerIlyThrIserIeuPro 199
601 TCTGAANAATGCTGCAACATCGAACAACATGAATTTGGCGCCGCGATCC 650
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 SerSerSerIleIserIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 216
651 TGCGGCTTGAAGCGGACGACATTCATTTCAATTCGAGCGCGGCGGCA 700
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 oAlaIgluIleuAlaIgluIThrIaIaIaIaIaIaIaIaIaIaIaIaIa 233
701 ATAAACCGGTGAGACCAATTCATTAAGATGATTAATTCATTTGCGCT 750
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 IuAsnValAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 249
751 TTGTTTGAACAGCGCGCTGTAACAGCGGCGGATTTGCCCTAGTGG 800
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 IeuPheIeuThrIgluIleuItyrThrAspArgValIaIaIaIaIaIaIa 266
801 TTCTCAAGTACAAACCGCGCTCTTGGTACCGCTTTGGTGGTGGCAAG 850
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 yProValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 283
851 TATCGCAATTTACTGCGGCGGATTTGTTGACACAGCAACACCGCGAT 900
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 eutIuAspIeuThrIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 299
901 TCCGCTTGGATTTGAACGCGCGGATTTACACAGCGCGCGACGATTT 950
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 SerGlySerValIleuSerGlyThrIaIaIaIaIaIaIaIaIaIaIa 316
951 GCGACGCTTACCAATCAATTCCTGATTCGATTCGAAGAAGCGCGCAAG 1000
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 uGIaTygTyHsIgluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 333
1001 AGCTGTTCGCGGCTGTTGCGCGCGACCGCAAAATCTCATCACGCT 1050
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
333 IuIeuPheIgluItyrPalaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 349
1051 ACAACCTGCGCGCATTTCTGAAAAACAACTTCAAGTTCAACACAGC 1100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 SerPheIeuIgluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 366
1101 CGTCAACGCGCGGACCGCGCATGCGGATTTGATGATGATGATGATG 1150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rThaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 383
1151 TGATGCGCTTGAATATCTGCGCGCGCGCTGTTTGGCGATTTATGCTC 1200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 aIeMetProIeuaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 399
1201 GCGGATACGCAACAGCGCGGATTTGGTGGTGGTGGTGGTGGTGGT 1250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 GlYAspThrAspSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 416
1251 AGACGCTGCTTGTGACGCTTGTGCGCGCGCGGCAAAATAGATAGAGGC 1300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uAspIeuaIaIeucIysThrPheValIysProIgluIysItyrGluIty 433
1301 CGTGTTCGCAAAAGTCTGGAACCAATGAGAGGAGGAGG 1341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 IuIeuIeuArgIuIuIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 446

```

seq_name: SwissProt_40:NQRA_VIBAL

seq_documentation_block:

ID NQRA_VIBAL STANDARD; PRT; 446 AA.
 AC 05686;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)

DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
 DE (Na(+)-translocating NADH-quinone reductase subunit alpha) (Na(+)-
 DE translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
 DE subunit A) (NOR-1 subunit A).
 GN NOR OR NOR1.
 OS *Vibrio alginolyticus*.
 CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 CC NCBI_TaxID=663;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
 RC STRAIN-NCIMB 11038;
 RX MEDLINE=95104445; PubMed=7805867;
 RA Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;
 RT "Cloning and sequencing of four structural genes for the Na(+)-
 RT translocating NADH-ubiquinone oxidoreductase of *Vibrio*
 RT *alginolyticus*.";
 RL FEBS Lett. 356:333-338(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Hayashi M., Umemoto T., Sugiyama A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 1-10.
 RX MEDLINE=98149659; PubMed=9490015;
 RA Nakayama Y., Hayashi M., Umemoto T.;
 RT "Identification of six subunits constituting Na⁺-translocating NADH-
 RT quinone reductase from the marine *Vibrio alginolyticus*.";
 RL FEBS Lett. 422:240-242(1998).
 RN [4]
 RN SEQUENCE OF 1-9 AND 334-340.
 RX MEDLINE=95104444; PubMed=7805866;
 RA Hayashi M., Hirai K., Umemoto T.;
 RT "Cloning of the Na(+)-translocating NADH-quinone reductase gene from
 RT the marine bacterium *Vibrio alginolyticus* and the expression of the
 RT beta-subunit in *Escherichia coli*.";
 RL FEBS Lett. 356:330-332(1994).
 RN [5]
 RN INHIBITION OF ENZYMATIC ACTIVITY.
 RX MEDLINE=20016049; PubMed=10549856;
 RA Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Umemoto T.;
 RT "Inhibitor studies of a new antibiotic, korormicin, 2-n-heptyl-4-
 RT hydroxyquinoline N-oxide and Ag⁺ toward the Na⁺-translocating NADH-
 RT quinone reductase from the marine *Vibrio alginolyticus*.";
 RL Biol. Pharm. Bull. 22:1064-1067(1999).
 RN [6]
 RN REVIEW.
 RX MEDLINE=21145117; PubMed=11248187;
 RA Hayashi M., Nakayama Y., Umemoto T.;
 RT "Recent progress in the Na(+)-translocating NADH-quinone reductase
 RT from the marine *Vibrio alginolyticus*.";
 RL Biochim. Biophys. Acta 1505:37-44(2001).
 RN [7]
 RN REVIEW.
 RX MEDLINE=21145118; PubMed=11248188;
 RA Steuber J.;
 RT "Na(+)-translocation by bacterial NADH:quinone oxidoreductases: an
 RT extension to the complex-I family of primary redox pumps.";
 RL Biochim. Biophys. Acta 1505:45-56(2001).
 CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
 CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
 CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NOR1 TO NOR2
 CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
 CC UBISEMIOQUINONE TO UBIQUINOL.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
 CC UBIQUINOL + NA(+)(OUT).
 CC -1- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+) +
 CC PUMPING REGULATION AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.
 CC INHIBITED BY KORORMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE
 CC (HONO).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NOR1, NOR2, NOR3, NOR4, NOR5
 CC AND NOR6.
 CC -1- SIMILARITY: BELONGS TO THE NOR1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z37111; CAA85476.1; -;
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
 FT CONFLICT 337 337 W -> L (IN REF. 4).
 SQ SEQUENCE 446 AA; 48622 MW; 6D65ACAA53FE515C CRC64;

alignment_scores:
 Quality: 1468.50 Length: 447
 Ratio: 3.814 Gaps: 1
 Percent Similarity: 86.130 Percent Identity: 62.416

alignment_block:
 US-09-303-518D-125 x NOR1_VIBAL

Align seg 1/1 to: NOR1_VIBAL from: 1 to: 446

```

1 ATGATTAAATCAAAAAGGCTAAACCTGCCCGCGGACGACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 MetIethrIleIysIsgIyleuAspIleuProIleAlaIleuLeuGly 17
51 GCAAGCGGTTTACGACGGCGGCCCATTTACGGAAGTGGCTGTTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 rGlnValIleAsnAspGlyIysThrIleIysIysValAlaLeuLeuGly 34
101 AAGAAATTCGGCGGTATCGCCCTCATGAAGTCAAGGAAGCGCATGCC 150
||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 IuGIuIyValIGlyMeIArgProThMeIHisValArgValGlyAspGlu 50
151 GTCAAAAAGGCCAAGTGTGCTTGAAGACAAAAGAACCGGCGGTGCT 200
||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 ValIysIysAlaIleuValIleuPheGluAlaIleuIysAsnProGlyVal 67
201 GTTTACTGCGCGGCTTCAGGCAAAATCGCGGATTCACCGTGGCGA 250
||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 sPheIThrAlaProIleAlaGlyIysValIleGluValaIleuArgGlyAla 84
251 AGCGCGTACTTCAGTACGTGATGCGGTGAAGCAAGCAAGCAAGAAC 300
||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ySarGylIeuGlnSerValIleGluValAlaIleuIysGluGluVal 100
301 GAGTTTGAACGCTACGACCTGAAGCGCTGCAAAATTAAGCGGGAAGA 350
||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ThrPheAspIysPheGluAlaIleuIleuSerGlyLeuAspArgGluVal 117
351 AGTGCAGCGGCAACCTGATCCGATTCGGTGTGGAGCTGCGCGACCC 400
||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 ILeIyThrGlnIleuValaIleuSerIleuThrIleAlaIleuArgGly 134
401 GTCCGTTACGCAAAATTCGCGGTGATGCGGAGCGGCTTGCACATTC 450
||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 rProPheSerIysValProIleAlaIleuIleuSerThrIysAlaIleu 150
451 GTCAATGCGATGAGCAACCAATCGCTGCGGACCGCTACGATGATAT 500
||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValThrAlaMetAspThrAsnProIleuAlaIleuIysProGluIleu 167
501 CAAAGAAGCGCGGAGAGATTTCAAAGCGCGCTGTGGATTATGACCGTT 550
||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 eaSngIuGlnGlnIleuAlaPheIleAlaGlyLeuAspIleuSerAla 184
551 TGACCGAAGCAAAATTCATGTTTGAAGCAAGCTGGCGGACAGCTGGCG 600
||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 euThrIuGlyIysValIyValIyValcysIys...SerGlyThrSerLeuPro 199
```

```

seq_name: SwissProt_40:NORA_VIBCH
seq_documentation_block:
ID   NORA_VIBCH                STANDARD;           PRT;       446 AA.
AC   O9KPSI; O9X4Q3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE   Na(+)-translocating NQR subunit A (Na(+)-NQR subunit A) (NQR complex
DE   subunit A) (NQR-1 subunit A).
GN   NORA OR VC2295.
OS   Vibrio cholerae.

```

34 LuGlutyrValIGlyMetArgprroThrmethIIsValArgValIGlyAspGlu 50

```

151 GTCAAAAAAGCCAGTGTCTTTGAGACAAAAAAGATCCGGCGTGT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 ValLysLysAlaGlnIleuPheGlnAspLysLysAsnProGlyVal 67
201 GTTTACTGCGCGGCTTACAGGAAATCCCGCATTCACCGTGGCGAA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 spherSerProValSerGlyLysValAlaGlnLeuAsnArgLysAla 84
251 AGCGGCTACTTCAGTCAGTCGTGATGCGGTTGAAGGACAGCAATC 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 ysArgValLeuGlnSerValValIleGluValAlaGlyAspSerVal 100
301 GAGTTGAACGCTTACGACCTGAGACGCGTGCACAACTTAAGCGGAGA 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 ThrPheAspLysPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 117
351 AGTGGCGCGCAACCTGATCCATCCGCTTTGTGACCTGCGTGGCGACC 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 alLeuLysThrGlnLeuValAlaGlnSerGlyLeuThrPheAlaPheArgThr 134
401 GTCCGTTGACGAAATTCCTGCGCTGATGCGGACCGCTTGCCATCTTC 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 rGProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 150
451 GTCAATGCGATGAGACCAACATCCGCTGCGTGCACCACTTACGTCATTAT 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 ValThrAlaMetAspThrAsnProLeuAlaIleGluProThrValIle 167
501 CAAGACAGCGCGCGAGATTTCAACGCGCGCTGTGTTGATGAGCGCTT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 eAsnGlnGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAla 184
551 TGACGCAAGCGCAAAATCCATGTTTGTAGAGAGCTGCGCGCAGACGTGCG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 euThrThrGlyLysValTyValLysLys...LysGlyThrSerLeuPro 199
601 TCTGAAATGCTGCCACATCGAAACACATGATTCGGCGCGCGCATCC 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 ArgSerGlnGlnProAsnValAlaGlnLeuIleValAlaPheAspGlyProIleSpr 216
651 TGCCGTTGAGTGTGACGACGACATTCATTCATGACGCGCGTGGCGCGGA 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 cAlaLeuAlaGlnIleThrHisMetHisPheLeuThrProValSerAla 233
701 ATAAACCGTGTGACGACATTCATTCATGATTAATTCATTCATTCATTCAT 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 sPheIleValAlaThrSerIleAsnLysArgLysAspValIleAlaValGlyLys 249
751 TTGTTTGAACAGCGCGCTGTGACACGCGCGCGTGGCGCGCATGCTGG 800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 LeuPheLeuThrGlyLysLeuThrGlnAlaArgValAlaSerLeuAlaGln 266
801 TTCTAGTCAACAAACCGCGCTGTGACGCGCGTGGCGCGCATGCTGG 850
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 lProValValAlaAsnLysProArgLeuValAlaArgThrValMetLysAlaSerL 283
851 TATGCCAAATTAATGCGCGCGCATGTTGACACAGACACACGCGGTGATT 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 euGlnGlnLeuValAspSerGlnIleMetProGlyGlnValArgLysIle 299
901 TCCGGTGGTATTGACGCGCGCATTCACAGACGCGCGCGCATTAATT 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyrLe 316
951 GGGACGCTACACAAATTCGATTCGTTATGAGAGAGCGCGCAAG 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
316 uGlyArgGlyThrHisLeuGlnValSerValLeuArgGlnLysArgAspLys 333
1001 AGCTTTGCGGTGGTGGCGCGCGCAAGCGCAATTAATTCATTCATTCAT 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
333 lLeuPheGlyTyrPheAlaMetProGlyLysAsnLysPheSerValThrArg 349
1051 ACAACCCTCGCGCATTTCTGAAAAACAACACTTTCAAGTTCAACACAGC 1100

```

```

350 SerPheLeuLysHisLeuPheLysGlyGlnValTyrAsnMetThrThr 366
1101 CGTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGlyLys 383
1151 TGATGCCCTTGATATCTGCGCCACCGCTGTTTGGCGCATTAATCTGC 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuGlyAla 399
400 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGlnLeuAspLys 416
1251 AGACCTGCTGTTGACGCTGCTGCTGCGCGCGCAATTCGATTCGCGC 1300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
416 uAspLeuAlaLeuLysThrPheValLysProGlyLysTyrGlyTyrGly 433
1301 CGCTGTTGCGCAAGTGTGAAACCATTCAGAGAGAGAGC 1341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 lLeuLeuArgGlnLysLeuAspLysIleGlnLysGlnLys 446

seq_name: SwissProt_40:NORA_PSEAE
seq_documentation_block:
ID NORA_PSEAE STANDARD: PRT; 445 AA.
AC O9H2K6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+))-translocating NOR subunit A (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR PA2999.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF004724; AAC06387.1; -
KW Oxidoreductase; NAD; ubiquinone; Transport; Sodium transport;

```


1297 GGCCCGCTGTGGCAAGTGTGCAACCATTTGAGAGGAA 1338
 446 PheAlAlleValLysGluAlaLeuIleArgTYrAlaLysGlu 455

seq_name: SwissProt_40:NORA_CHLPN

seq_documentation_block:
 ID NORA_CHLPN STANDARD; PRT; 467 AA.

AC 092752;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable Na(+)-translocating NADH-quinone reductase subunit A
 DE (EC 1.6.5.-) (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A)
 DE (NOR complex subunit A) (NOR-1 subunit A).
 GN NORA OR NOR1 OR CPN0743 OR CPN002.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;

SEQUENCE FROM N.A.

RA MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).

SEQUENCE FROM N.A.

RA MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).

SEQUENCE FROM N.A.

RA MEDLINE=20330349; PubMed=10871362;
 RA Shiba M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and Cw1029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).

FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO

UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
 OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
 ARE PROBABLY INVOLVED IN THE SECOND STEP. THE CONVERSION OF
 UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY)

-1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +

UBIQUINOL + NA(+)(OUT).

-1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NORB, NQRC, NQRE

AND NQRF (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE NORA FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AEO01656; AAD18882.1; -;
 DR EMBL: AEO02164; AAF37899.1; -;
 DR EMBL: AF002547; BAA96950.1; -;
 DR TIGR: CP0002; -;
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;

KW Complete proteome.
 SO SEQUENCE 467 AA; 52266 MW; 84CE73278BA0BFD6 CRC64;

alignment_scores:
 Quality: 464.50 Length: 464
 Ratio: 1.585 Gaps: 15
 Percent Similarity: 63.147 Percent Identity: 30.172

alignment_block:

US-09-303-518D-125 x NORA_CHLPN

Align seg 1/1 to: NORA_CHLPN from: 1 to: 467

4 ATTAATAATCAAAAAAGCTTAACCTGCGGAGGAGAGCG...GA 50
 ||| :
 3 lIeThrValAsnArgGlyLeuAspLeuSerLeuInglySerProlySgl 19
 51 GCAGACCGCTTACGAC.....GGCCGGCCATTACCGAAGTCGGTTC 94
 :
 19 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIleAspLeu 35
 95 TTGGCGAAGATATGCGGATGCGGCGCCCGCATGAGAAAGTCAAGAGAGC 144
 :
 36ArgProPheGlnProLeuSerLeuLysValGluIngly 50
 145 GATGCCCGTCAAAAAAGCGCAAGTGCCTTTGAACACAAAAGATCGGG 194
 ||| :
 51 AspAlaValCysSerGlyAlaProIleAlaGluTyrLysHisPheProAs 67
 195 CGTGCTGTTACTGCGCGGCTTCAGCAAAATCGCGCATTCACCGTG 244
 :
 67 nThrTyrIleTherHisValSerGlyValAlaThrAlaIleArgAryg 84
 245 GCGAAAGCGCGTACTTCAGTCACTGCTGAT...GCCGTGAAGGCAAC 291
 ||| :
 84 LysAluLysArgSerLeuAspValIleLeuLysThrProGlyPro 100
 292 GACGAAATCGAGTTTGAACGCTGACGACGTAAGCGTGGCAACTAAG 341
 :
 101 ThrSerThrGluTyr.....ThrTyrAspLeuInThrLeu 113
 342 CGGCGAAGATGCGCGGCAACCGATTCAGTTCAGTTGTGACGCGC 391
 | :
 113 rArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuPheAlaLeu 130
 392 TGGCGACCGCTGCGTTCAGCAAAATTCGCGCTGATGCGGAG...CG 438
 :
 130 LeuSglArgProPheAsp...IleProAlaIleProThrGlnThrPro 145
 439 TTGGCATCTTGTCAATGCGATGACACCAATCCGCTGCGGACCC 488
 :
 146 ArgAspArgPheIleAsnLeuAlaAspAsnArgProPheThrProSer 162
 489 TACGTCATATC.....AAGAAGCGCGGAGAGAT 520
 | :
 162 ogLysHisLeuAlaLeuPheSerSerArgGluGluGlyPheTyrVal 179
 521 TCAAGCGCGCTGTGATTTGAGCGTTTGACGCAAGCAAAATCAT 570
 ||| :
 179 heValValGlyValArgAlaIleAlaLysLeuPheGlyLeuArgProHis 195
 571 GTTGTGAAGCAGCTGCGGAGCGCGTGTGAAGATGTCGCAACAT 620
 :
 196 lIleValPheArgAspArgGlyLeuThrLeuProIleGlnGlyLeuLysThr 212
 621 C...GAACATGAAATTCGCGCGCGGCGGATCGCTGCGTTGAAGTGA 667
 :
 212 eAlaHisLeuHisThrValSerGlyProPheProSerGlySerProSer 229
 668 CGGATCATTTATTCATCGAGCGGCGTGGCGGCAATAA...ACGTTGG 714
 ||| :
 229 lHisIleHisSerValAlaProIleThrAsnGluLysGluValValPhe 245

```

715 ACCATCAATATTCAGATGTATACATTCGCGGTTGTTGCAACAGG 764
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 ThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPheLeuYsgl 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
765 CCGCTCAGAACACGACGCGGATGATGCGCTGATGCTTCAAGTCAACA 814
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 YargIleuHisIsgIuInValThrAlaLeuAlaGlyThrAlaLeuYsgl 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 AACCG.....CGCCTCTGCGTACCGTTTGGGTGGCAAGATATCGCAA 858
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 erSerLeuArgArgItyrValIleThrThrIlysgIyAlaSerPheSer 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
859 ..ATTATCGGGCGAATGGTGTGACACAGACACCGCGGATTTCCGG 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 LeuIleAsnLeuAsnAspIleSerAspAsnAspThr...LeuIleSerG 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 TTCGGTATTGAACGCGCGGATTTACACAGCGCGCAGCAT...TATTTGG 952
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 YAspProLeuThrGlyArgLeuCysIlysgIyGluGluProPheLeuG 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
953 GACGTTACCAATCAGATTCCTTCCTTATCGAAGAAGCGCGACGAAGA 1002
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 LypPheArgAspHisSerIleSerValLeuHisAsnProThrLysArgGlu 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 CTGTTGCGGTGGTGGCGCGCGACCGCAATATCTCATCAGCGCTAC 1052
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 LeuPheSerPheLeuArgIleGlyLysAsnLysProThrPheThrIly 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1053 AACCTCGGCGCATTTCTGAAACAAACCTCTCAAGTTCAAC.....A 1096
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 TyrIleuSerGlyPhePheLysIlysgIyArgThrTyrThrAsnPro 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1097 CAGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 hAsnLeuHisIleGlyLysThrArgProIleIleAspThrAspIleTyr 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1147 CGCGGTATGCGCTTGGATATCTCGCGCGCGCGCGCGCGCGCGCG 1196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 LysValMetProMetArgIleProValValProLeuIleLysAlaVal 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1197 CGTCGCGCATACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 eThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheLeuGluVal 428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1247 AAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 LysGluAspPheAlaLeuProThrIleAspProSerLysThrGluMet 444
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1297 GGCCCGCTGTTGCGCGCAAGTGTGGAACCATTTGAGAAGAA 1338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 LeuThrIleValLysGluSerLeuIleGlyLysThrAlaLysGlu 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: SwissProt_40:NORA_CHLTR

```

```

seq_documentation_block:
ID NORA_CHLTR STANDARD; PRT; 465 AA.
AC 084639;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable Na(+)-translocating NADH-quinone reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE (NQR complex subunit A) (NQR-1 subunit A).
GN NORA OR NQR1 OR CT634.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/DW-3/CX;
RX MEDLINE=99000809; PubMed=9784136.
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

```

```

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC UBIQUINOL + NA(+)(OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC EMBL; AE001334; AAC68238.1; -.
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
CC Complete proteome.
CC KW
CC SEQUENCE 465 AA; 51757 MW; 303F88DC133AB998 CRC64;

```

alignment_scores:

Quality: 446.00 Length: 464
Ratio: 1.517 Gaps: 14
Percent Similarity: 63.362 Percent Identity: 28.233

alignment_block:

US-09-303-518d-125 x NORA_CHLTR ..

Align seg 1/1 to: NORA_CHLTR from: 1 to: 465

```

4 ATTAATATCAAAAAGGCTTAACCTGTCATCGCGCGCGCGAGAGA 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 IleValValSerArgIlyLeuAspLeuSerLeuIlyAlaProIlysgl 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 AGCCGTTTACGACGCG.....CGGCGCATTCAGCAAGTGGCGTTGC 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 uSerGlyPheCysGlyLysValAspProThrIlyrValSerValAspLeu. 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 TTGGCGAAGAAATATGCCGATGCGCGCGCGCGCGCGCGCGCGCG 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 ....ArgProPheAlaProLeuProLeuGlyValIlyValThrProGlu 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 GATCGCGCTCAAAAAGGCGCAAGTGTGTTGAAGACAAAGAAATCCGG 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 AspGluValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerI 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 YValPheIleThrSerProValAspGlyLysValValGluIleArgG 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 GCGAAAAGCGCGTCTTCATCGTGTGATTCGCTTGA...GGCAAC 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 LysIlyAspGlyAlaLeuLeuGluIleValIleLysLysLysProGlyIle 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 GACGAATCGATTTGACGCGCTACGACCTGGAAGCGCTGGCAACTTAG 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 SerGluThrIlyPheSer.....TyrAspLeuGlnSerLeuThr 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 CGCGCAAGAAAGTGGCGCGCAACCTGATTCATCGGTTGTGGATCGCG 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 rGlnLysAspLeuLeuGluValAlaPheLysIlysgIyLysLeuPheAlaLeuP 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 TGCGACCGCGCTCGCTTCAGCAAAATTCGCGCGCTGCATGCCGAG...CCG 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

130 helysglnarprophasp...ileroalaleuprothrglnserpro 145
      ::::: |||||::: |||||::: ::::: ||
439 TTGGCCATTTTCGTAATGGCAGACCAACCAATCCGCTG...GCTGC 482
      ::::: |||||::: |||||::: ::::: ||
146 ArgAspValPheIleasnIleAspAsnArgProPheThrProSerVa 162
      ::::: |||||::: |||||::: ::::: ||
483 CGACCCACGAGCATTAACAAGACCGCGGAGAT.....T 520
      ::::: |||||::: |||||::: ::::: ||
162 lGluLysHisLeuSerLeuPheSerLeuYsgIuAspGlyTYrIleP 179
      ::::: |||||::: |||||::: ::::: ||
521 TCAAGCGCGCGCTGTGATTGACCGCTTGACGACGCAAAATTCAT 570
      ::::: |||||::: |||||::: ::::: ||
179 heValValGluValGlnAlaIleAlaLysLeuPheGlyLeuLysProHis 195
      ::::: |||||::: |||||::: ::::: ||
571 GTTTGAAGCGACCTGGCGCAGACGCGCTGTGAATAATGCTGCCAAT 620
      ::::: |||||::: |||||::: ::::: ||
196 lIleIleSerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIl 212
      ::::: |||||::: |||||::: ::::: ||
621 C...GAACACATGAAATCGCGCGCGCATCTCGCGGTGAGTGCGCA 667
      ::::: |||||::: |||||::: ::::: ||
212 eAlaHisLeuHisThrIleAspGlyProPheProSerGlySerProSerT 229
      ::::: |||||::: |||||::: ::::: ||
668 CGCACATTCATTCATCGACGCGCTC...GGCGCAATAAACCGGTGC 714
      ::::: |||||::: |||||::: ::::: ||
229 hrHisIleHisIleAlaArgIleArgAsnGlnArgAspValAlaPhe 245
      ::::: |||||::: |||||::: ::::: ||
715 ACCATCAATATTCACAGATTAATTCACATTCGCGCTGTGTCGACAGC 764
      ::::: |||||::: |||||::: ::::: ||
246 ThrIleSerPheGlnGluValLeuSerIleGlyHisIleuPheLeuLys 262
      ::::: |||||::: |||||::: ::::: ||
765 CGCTGTGACACCGACGCGCTGATTCGCTAGGTGTCTCAAGTC.... 810
      ::::: |||||::: |||||::: ::::: ||
262 yrPheValIleuGlyGlnGlnIleValAlaLeuAlaGlySerAlaLeuProp 279
      ::::: |||||::: |||||::: ::::: ||
811 ..AACAACCGCGCGCTTCGCTGACCGCTTGGTGCGCAAAATTCGCA 858
      ::::: |||||::: |||||::: ::::: ||
279 roSerGlnArgLysTyLeuIleThrAlaLysGlyAlaSerPheSerasp 295
      ::::: |||||::: |||||::: ::::: ||
859 ATTACGCGCGGGAATGTTGACACAGACAAAC...CGCGATTTCCGG 905
      ::::: |||||::: |||||::: ::::: ||
296 LeuLeuProLysAspIlePheSerSerAspGluIleThrLeuIleSerIl 312
      ::::: |||||::: |||||::: ::::: ||
906 TTGCGATTTGAACGCGCGCATTCACAAAGCGCGCAGCAT...TATTGG 952
      ::::: |||||::: |||||::: ::::: ||
312 yAspProLeuThrGlyArgLeuLysLysGlnGlnAsnProCysLeuG 329
      ::::: |||||::: |||||::: ::::: ||
953 GACGCTACCAATACAGATTCGCTTATCGAAGAGCGCGCAGCAAGAG 1002
      ::::: |||||::: |||||::: ::::: ||
329 lYmeIaArgAspHisThrIleThrLeuLeuProAsnProLysThrArgGlu 345
      ::::: |||||::: |||||::: ::::: ||
1003 CMTTCGGCTGGGTTGGCGCGCGCAGCAAAATCTCATACCGCGTAC 1052
      ::::: |||||::: |||||::: ::::: ||
346 SerPheSerPheLeuArgLeuGlyTTrpAsnLysLeuThrValThrArgTh 362
      ::::: |||||::: |||||::: ::::: ||
1053 AACCCCTGGCCATTTCTGAAAAACA...CTTTCAGATTCACCA 1096
      ::::: |||||::: |||||::: ::::: ||
362 lTyLeuSerGlyPhePheLysArgLysArgValAlaPheMetLysPheCAspT 379
      ::::: |||||::: |||||::: ::::: ||
1097 CAGCCGCTAAGCGCGCGCAGCGCGCATGTCGCCGATTGACTTACAG 1146
      ::::: |||||::: |||||::: ::::: ||
379 hrAsnMetHisGlyLysLysArgProIleIleAspAlaGluIleTyctIu 395
      ::::: |||||::: |||||::: ::::: ||
1147 CGCGGATAGCCGCTGATATCTGCCACCGCTGTTGGCGCGCATTTAT 1196
      ::::: |||||::: |||||::: ::::: ||
396 ArgValSerAlaIleProValProValAlaLeuIleIleLysAlaLeuGl 412
      ::::: |||||::: |||||::: ::::: ||
1197 CGTCGCGCATACCGACGCGCGCAGCGCATGCTGTGGAATTTGAGCG 1246
      ::::: |||||::: |||||::: ::::: ||
412 uThrGlnAsnPheGlnGluAlaLysArgLeuGlyLeuLeuGlnValAlaP 429
      ::::: |||||::: |||||::: ::::: ||
1247 AAGAAGACCTGCTTGTGACGCTTCCTGCTCCGCGCAATTCAGCAATAC 1296
      ::::: |||||::: |||||::: ::::: ||

```

```

429 roGlnAspPheAlaLeuProThrPheIleAspProSerLysThrGluMet 445
1297 GCGCCGCTGTCGCAAGTGTGGAACCAATGAGAGGAA 1338
      ::::: |||||::: |||||::: ::::: ||
446 PheSerIleValLysGlnSerLeuLeuArgTyAlaLysGlu 459
seq_name: SwissProt_40:RNFC_VIBCH

```

```

seq_documentation_block:
ID RNFC_VIBCH STANDARD; PRT; 774 AA.
AC O9KT88;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR VC1015
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=66;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Retchlin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnA, rnB, rnC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AE004183; AAF94176.1; ALT_INIT.
DR TIGR: VC1015;
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51k.
DR Pfam: PF001512; Complex1_51k; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
KW
FT METAL 378 378 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 381 381 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 384 384 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 388 388 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 417 417 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 420 420 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 423 423 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 427 427 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 427 427 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 774 AA; 83500 MW; 44EB13EC304E3A4E CRC64;

```

```

alignment_scores:
Quality: 158.00 Length: 472
Ratio: 0.721 Gaps: 16

```

Percent Similarity: 46.398 Percent Identity: 20.127

alignment block:
US-09-303-518D-125 x RNFC_VIBCH ..

Align seg 1/1 to: RNFC_VIBCH from: 1 to: 774

```

52 CAAGCGGTTTACGAGCGCCGCCATTCAGCAAGCGCGTTGCTGGCGA 101
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
32 GlnProIleIleAsnAlaSerIleProAsnGluLeuValIleuProLeu 48
   : : : : : : : : : : : : : : : : : : : : : : : :
102 AGAATATCCCGGTATGCGCCCTCGATGAAGATCAAGAGGCGATGCCG 151
   : : : : : : : : : : : : : : : : : : : : : : : :
48 sGlnHisIleGlyLysAlaGlyAspLeuLeuValIysValIysAspG 65
   : : : : : : : : : : : : : : : : : : : : : : : :
152 TCAAAAAGGCCAAGTCTGTTTGAGCAAAAAGAACATCCGGCGGTG 201
   ||| ||||| ||| : : : : : : : : : : : : : : : :
65 aLLeuLysGlyGlnProLeuThrGlnTyrThrSerThrPheMetLeuPro 81
   : : : : : : : : : : : : : : : : : : : : : : : :
202 TTACTGCGCGGCTTCAGCAAAATCGCGGATTCACGCTGGCGAATA 251
   : : : ||||| : : : : : : : : : : : : : : : :
82 ILeHisAlaProThrSerGlyValIleSerAlaIle.....GluPr 95
   : : : : : : : : : : : : : : : : : : : : : : : :
252 GCGCGTACTCAG.....TCAGTCGTGATG 277
   ||| : : : : : : : : : : : : : : : : : : : : : :
95 oArgThrValAlaHisProSerGlyLeuSerGlyLeuGlySileValLeu 112
   : : : : : : : : : : : : : : : : : : : : : : : :
278 CCGTGGAAGGCAAGCAGAA.....ATCGACTT 306
   : : : : : : : : : : : : : : : : : : : : : : : :
112 hProAspGlnGlnGlnGlnTyrPheGluLeuGlnProGlnProAspPhe 128
   : : : : : : : : : : : : : : : : : : : : : : : :
307 GAAGCGTACGACACTGAAGCGGTGCAAACTTAAGCGCGCAAGAGTGC 356
   : : : : : : : : : : : : : : : : : : : : : : : :
129 GlnGlnLeuThrProGlnThrLeuLeuGluLeuIleArgGlnAlaGly 145
   : : : : : : : : : : : : : : : : : : : : : : : :
357 CCGCAACTGATCAATCCGGTTTGTGACTGCGCTGCGCAACCGTCCGT 406
   : : : : : : : : : : : : : : : : : : : : : : : :
145 eSerGlyMetGlyGlyAlaGlyPheProThrAla..... 156
   : : : : : : : : : : : : : : : : : : : : : : : :
407 TCAGCAAAATTCGTCGCGTCATGCCGAGCGCGTTCGCATCTGTCAT 456
   : : : : : : : : : : : : : : : : : : : : : : : :
157 ..LysLysLeuGlnSerGlyLeuSerArgThrGluIleLeuIleAsn 172
   : : : : : : : : : : : : : : : : : : : : : : : :
457 GCGATGACACCAATCCGCTGGCTGCCGACCCCTACGCTATTATCAAGA 506
   ||| : : : : : : : : : : : : : : : : : : : : : :
173 AlaAlaGlnCysGluProTyrIleThrAlaAspAspValIleuMetArg 189
   : : : : : : : : : : : : : : : : : : : : : : : :
507 ACCCGCCGAGATTCAACAGCGGCGCTGTGATGACCGCTTGACCG 556
   : : : : : : : : : : : : : : : : : : : : : : : :
189 nTyrAlaHisGluIleIleGlnGlyIleGluIleValIleHisIleLeu 206
   : : : : : : : : : : : : : : : : : : : : : : : :
557 AACGCAAAATCCATGTT..... 573
   : : : : : : : : : : : : : : : : : : : : : : : :
206 yProLysLeuThrIleIleGlyIleGluAspAsnLysProGluAlaVal 222
   : : : : : : : : : : : : : : : : : : : : : : : :
574 .....TTAGAGCAGCTGCGCAGACAGCTCCCTCT..... 603
   : : : : : : : : : : : : : : : : : : : : : : : :
223 AlaAlaLeuGlnGlnAlaAlaGlnAspLysProMetValIleArgVal 239
   : : : : : : : : : : : : : : : : : : : : : : : :
604 .....GAAATGCTG 613
   : : : : : : : : : : : : : : : : : : : : : : : :
239 eProThrLysTyrProSerGlyGlyLysGlnLeuIleLysIleLeu 256
   : : : : : : : : : : : : : : : : : : : : : : : :
614 CCAACATCGAAGACAGATTCGCGCGCCGACCTTCGCGTTTGAGT 663
   : : : : : : : : : : : : : : : : : : : : : : : :
256 hAsnLeuLeuValProLysGlyGlyIleProAlaAspIleGlyLeu... 271
   : : : : : : : : : : : : : : : : : : : : : : : :
664 GGCAGCGACATTCATTCATGAGCGCGGTGCGCGGCAATAAACCGGTG 713
   : : : : : : : : : : : : : : : : : : : : : : : :
272 .....MetValGlnAsnValGlySer..... 278
   : : : : : : : : : : : : : : : : : : : : : : : :
714 GACCATCATTTATCAAGATTAATACCTTGCGGCTTGTGTTGCAACAG 763
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

279 .....LeuGlnAlaIleAlaArgAlaIleValHisG 289
764 GCCGTCGAACACGAGCGCGTATGTCCTAGTGTTCTCAAGTCAAC 813
   ||| : : : : : : : : : : : : : : : : : : : : : :
289 LysGluProLeuIleArgArgValValThrLeuThrGlyAspGlyPheArg 305
   : : : : : : : : : : : : : : : : : : : : : : : :
814 AACCGCGCGCTTGGCTACCGCTTTGGGTGCGAAGATTCGCAAAATAC 863
   ||||| : : : : : : : : : : : : : : : : : : : : : :
306 LysProArgAsnValThrPalLeuLeuGlyThrProValGlnAlaLeu 322
   : : : : : : : : : : : : : : : : : : : : : : : :
864 TCGGGCGAATTGGTGCACAGACAAAC.....CGCGATTTCCG 904
   : : : : : : : : : : : : : : : : : : : : : : : :
322 uAsnGlnPheGlyTyrLysAlaAspLysLysLeuProArgLeuIleMetG 939
   : : : : : : : : : : : : : : : : : : : : : : : :
905 GTTCGGATTGAAGCGCGGATTCACAGCGCGCATTTATTTGGGA 954
   ||| : : : : : : : : : : : : : : : : : : : : : :
339 LysIlyProMetMetGlyPheThrLeuPro..... 348
   : : : : : : : : : : : : : : : : : : : : : : : :
955 CGCTACCAACATCAGATTCCTTCATGAAAGAGCGCGCAAGAGCT 1004
   ||| : : : : : : : : : : : : : : : : : : : : : :
349 .....HisAlaGlnValProIle.....ThrLysThrAl 358
   : : : : : : : : : : : : : : : : : : : : : : : :
1005 GTTCGGCTGGGTGCGCGCGCGAGCGGACAAATATCTCATCAGCGTACA 1054
   : : : : : : : : : : : : : : : : : : : : : : : :
358 aAsnCysIleLeuAlaPro.....ThrArgAsnG 368
   : : : : : : : : : : : : : : : : : : : : : : : :
1055 CCTCGGACCATTTCTGAAAAACAACACTCTCAATTCACACAGCGCTC 1104
   ||| : : : : : : : : : : : : : : : : : : : : : :
368 LysLeu.....ThrSer 371
   : : : : : : : : : : : : : : : : : : : : : : : :
1105 AACGCGGCGACCGCGCCATGTGTCGAGTTGACTTACGAGCGCTCAT 1154
   : : : : : : : : : : : : : : : : : : : : : : : :
372 SerAspAsnGlnMetAlaCysIleArgCysGlyGlnCysAlaGlnLac 388
   : : : : : : : : : : : : : : : : : : : : : : : :
1155 GCCCTGGATATCTGCGCCACCTGCTTTGGCGCATTAATCGTGGCGG 1204
   : : : : : : : : : : : : : : : : : : : : : : : :
388 sProValSerLeuLeuProGlnGlnLeuGlnThrPheAlaLysAlaGlnG 405
   : : : : : : : : : : : : : : : : : : : : : : : :
1205 ATACCGACAGCGCGCAGGCAATTTGGTTGCTTGAATTGACGAAGAAGC 1254
   : : : : : : : : : : : : : : : : : : : : : : : :
405 LysPheAspLys.....CysGlnGluLeuAspLysLeuAsp 416
   : : : : : : : : : : : : : : : : : : : : : : : :
1255 .....CTCGCTTTGTGACACTTCGCTCGCCCGCGCAATACGAATA 1295
   : : : : : : : : : : : : : : : : : : : : : : : :
417 CysIleGlnCysGlyAlaCysAlaIleValCysProSerGluLeu..... 431
   : : : : : : : : : : : : : : : : : : : : : : : :
1296 CGGCGCGCTGTCGCG 1311
   ||||| : : : : : : : : : : : : : : : : : : : : : :
432 .....ProLeuValGln 435
   : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: SwissProt_40:RNFC_RHOCA

seq_documentation_block:

ID RNFC_RHOCA STANDARD: PRT; 519 AA.

AC 052715: 008057: 052712:

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Electron transport complex protein nfc (Nitrogen fixation protein nfc).

DE nfc).

DE nfc).

GN RNFC.

OS Rhodobacter capsulatus (Rhodospirillum rubrum).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

OC Rhodobacter.

OX NCBI_TaxID=1061;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=B105;

RX MEDLINE=9408454; PubMed=8264535;

RA Schmeil M., Jahn A., Meyer zu Vilsendorf A., Hennecke S., Masepohl B.,

RA Schuppler M., Marxer M., Oelze J., Klipp W.;

RT Identification of a new class of nitrogen fixation genes in

RT Rhodobacter capsulatus: a putative membrane complex involved in


```

RT electron transport to nitrogense."
RL Mol. Gen. Genet. 241:602-615(1993).
RN [2]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RP STRAIN-ATCC 33303 / B10;
RA MEDLINE-98151232; PubMed-9492268;
RX Jouanneau Y., Jeong H.-S., Hugo N., Meyer C., Willison J.C.;
RT "Overexpression in Escherichia coli of the rnf genes from Rhodobacter
RT capsulatus -- characterization of two membrane-bound iron-sulfur
RT proteins."
RL Eur. J. Biochem. 251:54-64(1998).
RN [3]
RP SUBCELLULAR LOCATION.
RC STRAIN-SBI003 / St. Louis;
RA MEDLINE-97299785; PubMed-9154934;
RX Kumagai H., Fujiwara T., Matsubara H., Saeki K.;
RT "Membrane localization, topology, and mutual stabilization of the
RT rnfBC gene products in Rhodobacter capsulatus and implications for a
RL Biochemistry 36:5509-5521(1997).
CC -1- FUNCTION: Required for nitrogen fixation. May be part of a
CC transbrane complex functioning as an intermediate in the electron
CC transport to nitrogenase. Stabilizes rnfB.
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG.
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (Probable).
CC -1- INDUCTION: Expression is reduced under iron-limiting conditions.
CC -1- SIMILARITY: BELONGS TO THE 4Fe4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X72888; CAAS139.1; .
DR EMBL, Y11913; CAAT72670.1; .
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1.51K.
DR Pfam: PF01512; Complex1.51K; 1.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 2.
KW Nitrogen fixation; Electron transport; Iron-sulfur; 4Fe-4S;
KW Inner membrane.
FT METAL 381
FT METAL 384
FT METAL 387
FT METAL 391
FT METAL 420
FT METAL 423
FT METAL 426
FT METAL 430
FT METAL 498
FT DOMAIN 503
FO SEQUENCE 519 AA; 55587 MW; B5451EB2A3FR6026 CRC64;

alignment_scores:
Quality: 150.00 Length: 454
Ratio: 0.676 Gaps: 18
Percent Similarity: 48.899 Percent Identity: 19.604

alignment_block:
US-09-303-518D-125 x RNF_C_RHOCA ..
Align seg 1/1 to: RNF_C_RHOCA from: 1 to: 519

```

```

120 CCCCIGATGAAAGTCAAGAGAGGCGCATGCCCTGCACAAAAGGCGCAATGCG 159
    : : : : : ||||| : : : ||| ||| ||||| : : :
56 uAlaGluPrOIlleValLyLysAtgAspAspLeuValLeuLysGlyGlnLeuI 73
170 TGTGTGAAGACAAAAGAAATCCGGGCGTGCTGTATTCTAGCGCGCGGTCA 219
    : : : : : : : : : : : : : : : ||||| : : :
73 LeAlaLysAlaArgGlyProLeuSerAlaAsnLeHisAlaProIleHisSer 89
220 GCCAAATGCGCGCGCATTT ..... CACGTGGCGGAAAA 251
    ||| : : : : : ||| : :
90 GlyArgValAlleAlaValAlcLynHisPheValAlaProHisAlaSerGlyLe 106
252 GCGCGATTCATGCAAGCGTGATGATTCGCGTGAAGCAAGCGAGAA... 297
    ||| : : : : : ||| : : : : : : : : : : :
106 uProVal...ProIleIleThrIleArgProAspGlyGlnAspLysIleTrrG 122
298 ..ATGAGATTGAACGCTACGCAACCTGAAGGCGCTGGCAAACTTAAACGGC 345
    : : : : : ||| ||||| : :
122 LPProHisLeuProArgLeuArgProGln.....AsnAlaAlaPro 135
346 GAGAAGTCCGCGCGCAACCTGATTCGAAATCCGGTTGTGG..... 384
    ||||| : : : : : : : : : : :
136 GluGluIleAlaIaGlnValAlaAlaIaGlyIleValaGlyMetGlyLeI 152
385 .....ACTGCGCTGCGGACCGCGCTGCATGCAGCAAAATTCCTG 421
    : : : : : : : : : : : : : : :
152 yAlaThrPheProSerAlaValLysLeuAsnLeuArgAlaIalysTyAspL 169
422 CCGTGATGCCGAGCGCTTCCGCATCTCGTCAATCCGATTCGACACACCAT 471
    : : : : : : : : : : : ||||| : : : : :
169 euThr.....ThrLeuIleLeAsnGlyAlaGlyGlyGln 180
472 CCGTGCGTCCGCGACCGCTACGCTCATTTCAAGAAGCGCGCGAGATTT 521
    ||| : : : : : : : : : : : ||||| : : : : :
181 ProTyrLeuThrCysAspAspArgArgMetArgGluArgAlaGlnGlnI 197
522 CAACGCGGCGCTGTGTGATTGAGCGCGTTGACCGACGCGCAAAATTCATG 571
    : : : : : ||||| : : : : : : : : : : :
197 eAlaSpolYIleGlyIleMetAlaArgAlaLeu..GlyValLysGlnVal 213
572 TTTGTA..... 577
    ||||| : : : : : : : : : : :
214 PheValAlaIleGluSerAsnLysProGlnAlaIleGlnAlaMetThrArg 230
578 .....AGGACGCTGGCGGACAGCTGCCGCTGCAAAATGCTGCCACAT 620
    : : : : : ||| ||||| : : : : : : : : : : :
230 gTyrAsnArgAlaLeuGluYIlyThrPheLysLeHisValValProIleHis 247
621 CGAAACACATCAATATTCGCGCGCGCGCATCCGCGCGTGTGAATGCGACGC 670
    : : : : : ||| : : : : : : : : : : :
247 InTyProMetGlySerGluLysHisLeuValLysMetIleThrGlyGln 263
671 ACATTCATTCATGAGCGCGTGGCGCGCAATAAACCGTGGACATC 720
    : : : : : : : : : : : ||| : : : |||
664 GluThrProAlaArgAlaLeuThrAlaLysPheu..GlyAlaValAlaHisA 280
721 AATTAACAAGATGTAATTAACATGCGCGCTGTGTTGGACAGCGCGCT 770
    : : : : : : : : : : : : : : : ||| : : :
280 snIleAlaThrAlaHis..AlaValHisLeuAlaValaArgTyGlyGluPr 296
771 GAACACCGAGCGCGTGAATGCGCTAGGTGTCTTCMACTCAACAAACGCG 820
    ||| : : : : : : : : : : :
296 OleuIleAlaArgThrValaThrValSerGlnIleGlyLeuIleArgArgPro 313
821 GCGCTTGGCGACGCTTTGGCGCGCAAGATGATGCGCAATTAACCTGCG... 867
    ||||| : : : : : : : : : : : |||
313 LaAsnLeuArgValLeuIleGlyThrProValSerGluIleLeAlaHis 329
868 ...GCGCAATTTGTTGACACAGACAGACCGCGTGAATTCGCGGTGGGATTT 914
    ||| : : : : : : : : : : : ||||| : : : : :
330 CysGlyLysIlyPheThrGlnGluProAspArgLeuLeuLeuGlyGlyProMe 346
1215 GAACGCGCGCATTCACACAGGCGCGGACGATTAATTTGGAGCGCTAACACA 964

```



```

163 1ueInlyVallylsleuilelleasnglyalaglucysgluprotyr 179
478 GGTGGCGACCCGACGTCATTCACAAAGACCCGCGAGATTCAACG 527
180 1lethCysaspsarlglemelarglyuarglalaasplullelely 196
528 CGGCCGTGTGATTCAGCGTTTGACCGACCAAAATCCATGTTTGA 577
196 sgllyleargyleuargtyrilleuethsprotlu..LysValValI 212
578 AGCGACCTGGCGACAGCGCCGCTC..... 603
212 1eal1legluaspsnlypsroglualalleaserAlalleargasnaIa 228
604 ...GAAATGTCCACATCGAACACATGATTCGGGGCGCCGACATCC 650
229 leugInglyAlaasnaplleserlleargyalleprothLysTyrP 245
651 TCCCGGTTTGAGTGGCAGCACATTCATTTCATCGAG.....C 688
245 oserglyalarhlyrsgInleuiletyrleuethrghlyleglualp 262
689 CGGTGGCGCCGAATAAACCGGTGACCATTAATTCAGATTAAT 738
262 roserglyluargserserlellyalValleuethlunasVal... 277
739 ACCATTGGCCGTTTGTTCACACAGCGCT.....CTGAC..... 774
278 ....Glythmethphealalleysaralalleleasnapglupr 292
775 ...ACGAGCGCGTATTGCCCTAGGTGTTCTCAGTCAACAACCGC 820
292 oleuilegluarValValthleuthrghlyasnllylealglulysg 309
821 GCCTTGTGACCGCTTTGGTGGCAAGTATCGCAAT.....ACT 864
309 lyAsrTyrTPalargleuglythPrroileserGlnlleuaserasp 325
865 GCGGCGAATTTGTTGACACAGACACACCGCGATTCCGGTGGGTATT 914
326 AlaGlytyrGlnpheaspllyshpheprolelealaglyglyProme 342
915 GAGCGCGCGATTACACAGCGCGCAGATTATTGGAGCTACACACA 964
342 tmetglyeu..... 345
965 ATCAATTTCCGTATCGAAGAGCGCGACCAAGAGCTGTGGCTGG 1014
346 ..gluLeuproasnlunasnaIaproValThrLysleuValasncysleu 361
1015 GTTGGCGCGACCGGACCAATACTTCATCAGCGGTACACACCTCGGCA 1064
362 leuAlaproaspyrlyleuglytyr..... 369
1065 TTTCTGAAAAAACTCTTCAAGTTCACACAGCGCTCAGCGCGCG 1114
370 .....AlagluProglunIaG 375
1115 ACCCGGCATGTGCCATTTGCTACTACGACCGCTGATCCCTTGAT 1164
375 luginAlaCysleargCyserserCysSerAspAlaCysProvalasn 391
1165 ATCCGTGCCACCGCTTTTGGCGGATTTATCTGCGCGCTACGACAG 1214
392 leuethProglInglInleuTytrPhealargserserlunaspshlysl 408
1215 CGCGCAGCATTTGGT.....TCCTTGAATTGGACGAAGACACC 1255
408 sserglyluInlyrAlaLeuLysasCysIleglu.....C 420
1256 TCCTTTGTGAGCTGTCTGCGCGCGGCAAAATACGAATACGCGCCCTG 1305

```

```

420 ysglylleCysAlaTyValCysPro.....SerHisIleProLeu 433
1306 TTGGCGC.....AAAGTGTGGAAACCATTTGAGAA 1334
434 1leGInTyrPheargInglulysAlaLysIletrpInlleuysgluTy 450
1335 GGAA 1338
450 sGIn 451

seq_name: SwissProt_40:HKRI_YEAST
seq_documentation_block:
ID HKRI_YEAST STANDARD: PRT: 1802 AA.
AC P41809;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Hansenula MRKII killer toxin-resistant protein 1 precursor.
GN HKRI OR YDR420W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YN 295;
RX MEDLINE=94156857; PubMed=8113191;
RA Kasahara S., Yamada H., Mio T., Shitatori Y., Miyamoto C.,
RA Yabe T., Nakajima T., Ichishima E., Furuchi Y.;
RT "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT overcomes the effects of HM-1 killer toxin, which inhibits
RT beta-glucan synthesis."
RL J. Bacteriol. 176:1488-1499(1994).
CC -1- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
CC PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
CC -1- SIMILARITY: SOME TO YEAST MSB2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, S69101; AAB30051.1; -.
DR SGD: S0002828; HKRI.
KW Glycoprotein; Transmembrane; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 1802
FT TRANSMEM 1486 1506
FT DOMAIN 23 1478
FT DOMAIN 453 788
FT REPEAT 453 480
FT REPEAT 481 508
FT REPEAT 509 536
FT REPEAT 537 564
FT REPEAT 565 592
FT REPEAT 593 620
FT REPEAT 621 648
FT REPEAT 649 676
FT REPEAT 677 704
FT REPEAT 705 732
FT REPEAT 733 760
FT REPEAT 761 788
FT CARBOHYD 24 24
FT CARBOHYD 1252 1252

```


295 CCGTCGTCCTTACAGCGCATCAGCTGAGTACGCGCTTTTCG 246
 868 LaserSerAlaGlnThrAla..... 874
 245 CCAGCGGATTCGCGGATTTTGCCTGAAGCGCGCAGTAACACAC 196
 875AlaSerHisSerSer 879
 195 GCCCGGATTC.....TTTTCGTCCTCAACAGCAGCTTGGC 161
 879 TGYGlyPheserGlyPheGlyGlyThrLeuThrThrSerThrAlaP 996
 160 CTTTTCGTCGTCGTCCTTCCTTCCTTCATTCGAGGCGCATACCG 111
 896 roLamrThrSerGlnProThrLeuThrPheSerSerThrValThrPro 912
 110 GCATATTCCTTCGCCACGACGACTTCGTAATGCGCGCGCTGCTA 61
 913 ThrPheSerThrLeuPheSerAlaSerAlaLeuProThrTy 929
 60 AACGCTTCCTCGCTCGCTCGCGCGGATGGC 30
 929 rProGlyAlaSerSerGlnProThrPheGly 939

seq_name: SwissProt_40:RNFC_PSEAE

seq_documentation_block: PRT: 774 AA.

ID RNFC_PSEAE STANDARD: PRT: 774 AA.
 AC 09HYB8;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 GN Electron transport complex protein rnfC.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RP SEQUENCE FROM N.A.
 RP STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wachman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 RT opportunistic pathogen.
 RL Nature 406:959-964(2000).
 CC - FUNCTION: May be part of a membrane complex involved in electron
 CC transport (By similarity).
 CC - COFACTOR: Binds 2 4Fe-4S clusters (Potential).
 CC - SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
 CC rnfD, rnfE and rnfG (By similarity).
 CC - SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC - SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
 CC RNFC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE004770; AAC06879.1;
 DR InterPro: IPR001450; 4Fe4S_Ferredoxin.
 DR InterPro: IPR001949; Complex1_51K.
 DR Pfam: PF01512; Complex1_51K; 1.
 DR Pfam: PF00037; fer4; 2.

DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
 KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
 KM Complete proteome.
 FT METAL 369 369 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 372 372 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 375 375 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 379 379 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 408 408 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 411 411 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 414 414 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 418 418 IRON-SULFUR (4FE-4S) (POTENTIAL).
 SQ SEQUENCE 774 AA; 82359 MW; C82B06F4D2E09DB6 CRC64;

alignment_scores:
 Quality: 129.00 Length: 488
 Ratio: 0.578 Gaps: 23
 Percent Similarity: 45.697 Percent Identity: 21.516

alignment_block:
 US-09-303-518D-125 x RNFC_PSEAE ..

Align seg 1/1 to: RNFC_PSEAE from: 1 to: 774

28 CCGCCATCGCGCGCGACACCGGACGACCGGCGCGCCAT 77
 |||||
 23 LeuProIleGlnGlnAlaProLeuAlaGlyTrpIleValPro..... 37
 78 TACCGAAGTCGCGTCTTCGCGCAAGAAAT...GCCGTAATGCCCCCT 124
 |||||
 38LeuGlyGlnHisIleGlyAlaProAlaValProC 49
 125 CGATGAATCAAGAGAGCGATGCCGTCGCAAAAAGGCCAAGTGTCTT 174
 :|||
 49 yValGluVal.....GlyGlnAlaValLeuGlyGlyThrIleAla 63
 175 GAAGCAAAAAGAAATCCGCGCGCTGCTTACTGCGCGCTTCAGGCA 224
 :|||
 64 LeuProASPglYThrValSerAlaIleuHisAlaProThrSerClyTh 80
 225 AATCGCGCGGAT.....CACCGCGCAAAAAGCGCGTAC 259
 :|||
 80 yValValAlaIleGlyAlaHisProGlyProHisAlaSerGlyLeuProA 97
 260 TTCAGTCAGTCGTGATGCCGCTGAAGCAAGCAGGAA..... 297
 :|||
 97 LeuAlaIleValIleAlaSerAspGlyLeuGluArgTrpThrIleu 113
 298ATCGAGTTGAAGCTAAGCAGCAGCTGAAGCGGCAACTT 338
 :|||
 114 HisProCysProAspPheArgAlaGluSerProLeuAlaIleu..... 128
 339 AACGCGCGAAGAGTCGCGCC.....AACGTATCCAAATCCG 376
 :|||
 129GluArgIleArgAlaAlaGlyIleGlyLeuGlyGlyAlaG 143
 377 GTTGTGAGACTGGC.....CTGCGCAGCCGTCCTGCAAAATTCCT 420
 :|||
 143 LysPheProThrAlaAlaLysLeuAlaAlaArgProAlaGluLysIle... 158
 421 GCCGTGATGCCGAGCGGTCGCAATCTTGCAATGCGATGACCAACA 470
 :|||
 159HisThrLeuValValAsnGlyAlaGluCysG 169
 471 TCGCGTCGCGCGCGACCTAAGCTCATTTTCAAGAGCGCGAGGAT 520
 :|||
 169 uProTyrIleSerAlaAspAspLeuLeuLysGluArgAlaThrGlnV 186
 521 TCAACGCGCGCTGTGATTCAGCGCT..... 549
 :|||
 186 allLeuGlyGlyIleAspIleLeuValGlnIleuGlyProGluGluVal 202
 :|||
 550TTGACCGAAGCAAAAATTCATGTTTGAAGGACAGCTGGCC 590


```

203 LeuVal1GlyLeuAspPylsProGluAla1LeuAlaLeuGly1 219
591 A.....GACGTCCCTCTAATAATGTCGCAACATGGAACATG 631
219 AlaLeuGlyLeuArgProTyrArgLeuAlaLeuProThrArgTyrP 236
632 AATTGGCGCGCCGCAT..... 648
236 roserGlyGlyLeuArgGlnLeu1LeuLeuThrGlyArgGluVal 252
649 CTTGCC...GGTTGAGTGGCAGCAGCATTTTCATGACCGCGTGG 695
253 ProAlaAspGlyLeuProAlaAsp1LeuGlyLeuGlyGlnAsnValG1 269
696 CGCGAATTAACCGTGTGACCATCATATATCAAGTGAATGATTCATTG 745
269 YThrLeuAlaAlaVal.....HisAspAlaValValLeuG 281
746 GCGGTTGTTGGACAGCGCGCTGACACCGCGGATTCGCTA 795
281 YArgProLeuIleSer.....ArgIleThrLeu 291
796 GGGGTTCACAGTCAACAACCGCGCTTGGTACCGTTTGGGTGC 845
292 AlaGlyGlyAlaLeuGlnArgProMetAsnValGluAlaLeu1GlyTh 308
846 GAAAGTATGCAATTAATCTGCG.....GGCAATGCG 877
308 ProValHisGluLeuLeuAlaPheAlaGlyLeuAlaGlyValArgLeu 324
878 TTGACACAGACAAACCGGTGATTCGCGTTCGATTTGACGCGCGATT 927
325 .....GluArgValLeuMetGlyGly..... 331
928 ACACAAGCGCGCAGATTATTGGAGCGATACCAATCAGATTCCGT 977
331 ..... 331
978 TATCAGAAAGCGCGCAGCAAGACCTGTTCGGTGGGTCGCGCAGC 1027
332 .....PrometMetGlyPheAlaLeuProAsp 340
1028 CGGACAATATCTCCATCAACGCGGTACAAACCTTCGCGCAATTCGAA 1077
341 .....LeuSerVal 343
1078 AACTCTTCAAGTTCAACACAGACCGGTCAACGCGCGCAGCGCATGTT 1127
344 ProLeuIleGlyThrCysAsnGlyLeuAlaGlyAspAlaThrGluLe 360
1128 GCGG.....ATTGTAATCTTACGAGC 1147
360 ProGluProValProAlaMetProCys1LeuArgCysGlyAspCysAlaG 377
1148 GCGTATGCGCTGTGATATCTGCCCAACCGCTTGGCGCGCATTAATC 1197
377 InValCysProValSerLeuLeuPro1L...GlnLeuHisPheAla 392
1198 GTGCGCGATACCGACAGCGCGCAGCATTTGGTTCGTAATGAGCA 1247
393 LeuGlyAsp...GlnHisGlnGlnLeuLeuAlaHisAsnLeuPheAsp 408
1248 AGAAGACGCTGCTTGTGACAGCTTCGTCGCGCGCAATACGATACG 1297
408 s1LeuGlyCysGlyAlaCysAlaTyrValCysProSerSerIle..... 422
1298 GCGCGCTGTTCGCG 1311
423 ..ProLeuValGln 426
seq_name: SwissProt_40:RNFC_BUCAI

```

```

seq_documentation_block:
ID RNFC_BUCAI STANDARD: PRT; 473 AA.
AC P57215;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Election transport complex protein rnfc.
GN RNFC OR B0115.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxId=118093;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=TOROYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -! FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -! COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -! SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -! SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -! SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.slb.ch)
-----
DR EMBL: AP001118; BAB12833.1; -
DR InterPro: IPR001450; 4Fe4S_Ferredoxin.
DR InterPro: IPR001949; Complex1_51k.
DR Pfam: PF001512; Complex1_51k; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 337
FT METAL 337
FT METAL 340
FT METAL 340 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 343 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 347 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 347 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 376 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 379 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 382 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 382 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 386 IRON-SULFUR (4Fe-4S) (POTENTIAL).
SQ SEQUENCE 473 AA; 54563 MW; 04B64102F1315AEE CRC64;
-----
alignment_scores:
Quality: 126.00 Length: 449
Ratio: 0.575 Gaps: 22
Percent Similarity: 48.775 Percent Identity: 20.045
alignment_block:
US-09-303-518D-125 x RNFC_BUCAI
Align seg 1/1 to: RNFC_BUCAI from: 1 to: 473
100 GAAAGATATGCGGTATGCGCCCTCGATGAAAGTCAAGAGCGGATGC 149
.....
13 AspAspTyrHisAsnValGly.....LeuArgValGlyIleAsnGluGly 27
150 GCTCAAAAAGGCCCA...GTGCTGTTGAAGACCAAAAAGATCCGGGCG 196
||| .....
27 sValLeuArgGlyGlnProLeuIlePheSerAspAspPheAsn.....V 42

```


US-09-303-518D-125 x SGS3_DROER ..

Align seg 1/1 to: SGS3_DROER from: 1 to: 328

```

29 TGGCCATCGCGGAGACCGGAGCAAGCCGTTACGAGCGCCG..... 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 CysProLysArgThr...ThrProLysProCysThrThrAlaArgProTh 43
74 .....CCATTACGACGAGCGGTGCGTGGCGAAGATATGCGCGTAT 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 rGysAlaProValThrThrThrCysArgProProThrThrThrArgC 60
117 GCGCCCTCGATGAAAGTCAAGAGGCGCATCGTCAAAAAAGGCCAAG 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 ysrProProThrThrThrThrArgCysProProThrThrArgProAla 76
167 TGGTGTTCAGACAAAGATCCGGCGGTGCTTACTGCGCCG..... 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 CysThrAlaThrThrLysArgProThrAlaArgProThrThrArgArg 93
215 ...CTTCAGGCAAAATCGCGGATTCACCGTGGCGAAGCGGCTACT 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 rThrValArgAlaThrThrLysArgAlaThrThrArgArgThrThrLys 110
261 TCAGTCAGTCG...TGATTGCCGTTGAAGGACAGACAGAA...TCGAGT 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 rGAlaThrThrArgArgThrThrValArgAlaThrThrLysArgAla 126
305 TGAAGCGTACGACACCTGAAAGCGCTGGCAACT...TAAGCGCGAAGAA 351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 ThrArgArgThrThrThrLysArgAlaProThrArgArgThrThrLys 143
352 GTGCGCCCGACCTGATCCATCCGCTT...TGAGACTGCGCTGCG 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 sArgAlaThrThrArgArgAsnProThrArgArgThrThrThrArgArg 160
396 CACCGCGTCCGTCACGAAATTCCTGCGCGTGCATCCGACCGCTGGCCA 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 lAprThrLysArgAlaThrThrLysArgAlaThrThrArgArgAsnPro 176
446 TCT.....TCGTCAATGCGATGACACCAATCCGC..... 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 ThrLysArgLysThrThrArgArgThrThrValArgAlaThrLysThr 193
476 .....TGGTGGCGACCGCTACGCTATTAACAAGA 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 rLysArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLys 210
507 AGCGCGCG.....AGGATTCACAAAGCGCGCTGTGTATGAGCGCTT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 rGAlaProThrLysArgValThrThrLysArgAla..... 221
551 TGACCGAAGCAAAATCCATGTTGTAAAGCAGCGTGGCGAGAGTGGCG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 ...ProThrLysArgAlaThrThrLysArgAlaProThrLysArgAla 237
601 TCTGAAATGCTGCCACATGAAACACATCAATTCGGCGCGCGCATCC 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 rThrLysArgAlaProThrLysArgAlaThrThrLysArgAlaProThr 254
651 TGCGCGTTGATGGCAGCG..... 670
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ysaArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLysArg 270
671 ...ACATTCAATTCATCGAGCGCGTGGCGAATAAAGCGTGGAGC 717
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 AlaThrAlaArgProThrSerLysProCysLysLysProCysGlyPro 287
718 A 718
287 o 287

```

seq_name: SwissProt_40:YM96_YEAST

seq_documentation_block:

```

ID YM96_YEAST STANDARD; PRT: 1140 AA.
AC 004893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRES-FET4 intergenic region.
GN YMR317W OR YM9924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SZ88C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z5141; CAA90835.1; -.
CC DR SGD; S0004936; YMR317W.
CC KW Hypothetical protein; Repeat.
CC
CC SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

```

alignment_scores:

```

Quality: 124.00 Length: 466
Ratio: 0.588 Gaps: 21
Percent Similarity: 45.279 Percent Identity: 23.176

```

alignment_block:

US-09-303-518D-125/rev x YM96_YEAST

Align seg 1/1 to: YM96_YEAST from: 1 to: 1140

```

1331 TCAATGTTCCAGACCTTGGCCACAGCGGCGCTATTCGATTGGCC 1282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 SerSerThrSerSerValSerSerGluAlaProSerSer..... 399
1281 CGGGCAGACGACGCTGCACAAAGCGAGTCTTCTGTCACATTCGACG 1232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 .....ThrSerSerValSerSerGluAla 408
1231 AACCCATGCTCGCGCGTGGTATCGCCAGATTAATCGCCGCAAA 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 lAprSerSerThrSerSerValSerSerGluAlaProSerThrLys 424
1181 AGCAGGTGGGACGATATCCAGGCGATCAGCGCTGCTAGTACCAAT 1132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 SerSerVal.....MetSerSerGluValSerSerAla.....Th 436
1131 CGGACCATGCGCGGTGCGCGCGCTTGACGCGCTGTGTAACCTGAGA 1082
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 rSerSerLeuValSerSerGluAlaProSerAlaLeuSerSerLeuAla 453
1081 GT.....TTGTTTTCAGGAAATGCGCAGGAGGTGTACGCTGATG... 1041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 eSerArgLeuPheSerSerLysAsnThrSerValThrSerThrLeuVal 469
1040 .....GAGTATTTCGCGGCTGGCGGCGCAACCCAGCGACGCTCTT 997
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 AlaThrGluAlaSerSerValThrSerSerLeuArgProSerSerGlu 486
996 GCTGCGCGCTTCTTCGATAACGAAATCGATTGATGTCGCTCCCAAT 947

```

```

486 rLeuAlaSerAsnSerIleIleGlu..... 494
494 ..... 494
896 AATGTCGGCGCCTTGTGTATCGCGCGTTTCATACCGAACGGAAATC 897
494 ..... 494
896 ACGGGTGTCTGTGTACACCAATTCGCCGACATAATTCGATCTT 847
495 SerSerLeuSerThrGlyTyrAsnSerThrValSerThrThrSerAl 511
846 CGCACCCAAACGGTACGCAAGCGCGGTTTGTACTTGAGAACAC 797
511 AlaSerSerThrLeuGlySerLys..... 519
796 CTAGGGCAATCACGCGCTGTTCAGACGCGCTTTCGAAACAAACG 747
520 ..... ValSerSerSerAsnSerAlaMetAlaThrSerLysThr 532
746 CCAATGTAATACATCTTGAATTAATGATGTCACACGGTTTATTCG 697
533 SerSer.....ThrSerAspLeuSerLysSerSerValIlePheG 547
696 CGCGACGCGCTCGATGAAATGATGCGTGCACCTCAACCGCAGGAT 647
547 YAsnSerSerThrVal.....T 553
646 GCGGCGCCGCAATTCATGTTTCGATGTTGCAGCA..... 609
553 hrThrSerProSerAlaSerIleSerLeuThrAlaSerProLeuProSer 569
608 .....TTTTCAGACGCGCTGTCCCGCAGCTCCCTTACAAACATGATTT 562
570 ValThrSerAspIleThrSerSerGluAlaSerIleSerSerAsnLe 586
561 CGGTTGCGTCAACGCGCTCAATACCAACGCGCGTTTGAATCCTCG 512
586 uAlaSerSerSerAlaProSerAspAsnAsnSerThrIleAlaSer.... 601
511 CGGCTCTTTGATTAATGACC.....GTGAGGTCGCGCA 480
602 ..AlaSerLeuIleValThrLysThrLysAsnSerValValSerSerIle 617
479 GCCACGCGATGTGTGTCATGCGATGACGAAGATGCGACGCGTCGCG 430
618 ValSerSerIleThrSerSerGluThrAsnGluSerAsnLeuAlaTh 634
429 ATCGACGCGCAGGATTTTGTGACGACGAGCGGTGCGCAGCGACATC 380
634 rSerSerThrSerLeuLeuSerAsn..... 642
379 AACCGGATTTGATCAGTTGGGGGCGACTTGTGCGCGTTAAGTTGCG 330
643 .....LysAlaThrAlaArgSerLeuSerThrSer 652
329 ACGCGTTCAGTGCCTAGCGTTCAAC.....TCGATTTGCTC 292
653 AsnAlaThrSerAla.....SerAsnValProThrGlyThrPheSerSe 667
291 GTTGCCTTCA..ACGCAATCAGCATGATGAACTACGCGCTTTTGGC 245
667 rMetSerSerIleThrSerVal..... 674
244 CACGGTGAATCGCGCGATTTTGCCTGAACGCGCGCAGTAACACACG 195
675 .....IleThr 676
194 CCGGATTTCTTTTGTCTTCA.....AACACACTTGGCC 160
677 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThr.... 691
159 TTTTTCGACGATCGCTTCTTACTTTATTCGAGGGGCGCATACCGG 110

```

```

692 .....ValValSerSerSerLeuAlaGlyTyrSerPheSerThrProG 706
109 CATATTCTTCGCAACAAACGCGACTTGCTATACCGCGCGCTGCTGA 60
706 lu...SerSerProThrThrSerThrLeuValThrSer..... 717
59 ACGGCTGTGCTGCGTGTCCGCGATGGGCGACGTTTACACCTTTTGG 12
718 GluAlaProSerThrValSerSerMetThrThrSerAlaProPheIle 733
seq_name: SwissProt_40:RNFC_ECO57
seq_documentation_block:
ID RNFC_ECO57 STANDARD; PRT; 740 AA.
AC P56324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfc.
GN RNFC OR Z2636 OR EGS2338.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Rose D.J., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Perna N.T., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Iln J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa G., Ogasawara N., Yasunaga T.,
RA Kudara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -!- SUBUNIT: Composed of at least six subunits: rnfa, rnfb, rnfc,
CC rnfd, rnfe and rnfg (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE005386; AAG56618.1; -
CC DR EMBL: AP002558; BAB35761.1; -
CC PROSITE: PS00198; 4FE4S-FERREDOXIN; 2.
CC KW Electron transport, Iron-sulfur, 4Fe-4S, Inner membrane.
CC FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: AE000258; AAC74701.1; -
 DR EMBL: D90806; BAA15384.1; -
 DR EMBL: D90807; BAA15391.1; -
 DR EMBL: D90808; BAA15414.1; -
 DR HSP: P00198; 1FCA.
 DR EcoGene; EG13935; rnfC.
 DR InterPro; IPR001450; 4Fe4s-ferredoxin.
 DR InterPro; IPR001949; Complex1_51K.
 DR Pfam; PF01512; Complex1_51K; 1.
 DR PROSITE; PS00198; 4Fe4s-FERREDOXIN; 2.
 DR Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
 KM Complete proteome.
 FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
 SEQUENCE 740 AA; 80171 MW; 381506475CD6E01F CRC64;

Alignment_scores:
 Quality: 115.00 Length: 322
 Ratio: 0.728 Gaps: 12
 Percent Similarity: 49.068 Percent Identity: 22.050

Alignment_block:
 US-09-303-518d-125 x RNFC_ECOLI ..

Align seg 1/1 to: RNFC_ECOLI from: 1 to: 740

100 GAGAGATATGCGCGGTATGCGCCCGTGAAGAGCAAGAGCGGATGC 149
 48 LysGlnHisIleGlyAlaGluGlyLeuGlyValSerValGlyAspGly 64
 150 CQTCAAAAAGGCGCAAGTCTGTGTAAGACAAAAGAAATCCGGCGG 199
 64 sValLeuArgGlyInProLeuThrArgGlyArgGlyLysMetLeuProV 81
 200 TGTCTACTGCGCGCGCTTCAGCAAAATCCGGCGAT..... 237
 81 AlHis...AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96
 238CAACCTGGGCAAAAGCGCTACTTCAGTCAGTCGATTCGCGT 281
 97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112
 282 TGAAGCAACGAC..... 294
 112 aAspArgGlyAspGlySTripIleProArgAspGlyThrAlaAspGlyArgT 129
 295GAAATCGAGTTTGAACGCTACGCACTGAAAGCGGTGCA 333
 129 hrArgSerArgGluGluLeuIleGluArgIleHisGlnPhenGlyAla 145
 334 AACTTAAGGCGGAGAGAGAGTGGCGGCAACTGATCCATCCGGTTGTG 383
 146 GlyLeuGlyGly.....AlaGlyPhePr 153
 384 GACTCGGTCGACACCGCGTTCGACGAAATTCGTCGTCGATGCGG 433
 153 ohnArgGlyAla.....LysLeuGlnGlyGlyGlyAspL 164
 434 AGCGTTCGCGCATCTTGTCAATGAGATGACACCAATCCGTCGCTGCC 483

164 ystIleGluThrLeuIleAlaAlaGluGlySerProGlyThr 180
 484 GACCCATCGGTCTATTCAAGAGCGCGCGAGATTTCAAACGCGGCT 533
 181 AlaAspAspArgLeuMetGlnAspGlyAlaAlaGlnValValGluIle 197
 534 GTTGATTTAGCGGTTTGACCGAA..... 558
 197 eArgIleLeuAlaHisIleLeuGlnProArgGluIleLeuIleGlyIle 214
 559CGCAAAATTCATGTTTGAAGCGACCTGGCGGAC 594
 214 IuAspAsnLysProGlnAlaIleSerMetLeuArgValAlaLeuAlaAsp 230
 595 GTCCGCGTGAATAATCGTCCGCAACATGCAACATGATTCGGCGCGCC 644
 231SerAsnAspIleSerLeuArgValIleProThrLys 242
 645 GCATCTGCGCGT.....TTGAGTGCACGCG 670
 242 sTryProSerGlyGlyAlaLysGlnLeuThrGlyIleLeuThrGlyLysG 259
 671 ACATTCATTTCAATGACCGCGCGCGCAATGAAACGCTGGACCATC 720
 259 InVal.....ProHisGlyGlyArgSerSerAspIleGlyVal 271
 721 AATTATCAAGATGTAATTAAC.....ATTGGCGGTTGTTTGCAC 761
 272 LeuMetGlnAsnValGlyThrAlaGlyAlaValLysArgAlaValIleAs 288
 762 AGCGCGTGTGACACCGAGCGCGTGAATGCGCTGAGTTCATGATCA 811
 288 rGlyGluProIleThrGluArgValValThrLeuThrGlyAlaIleAla 305
 812 ACAACCGCGCGCTTGGTACCGTTTGGTGGCAATATTCGCAAT 861
 305 IaaArgProGlyAsnValITrPalaArgLeuGlyThrProValArgHisLeu 321
 862ACTGCGGCAATGTTGACACAGACACGCGGATTCGCG 905
 322 LeuAsnAspAlaGlyPheCysProSerAlaAspGlnMetValIleMet 338
 906 TTCGATTTGACGCG 921
 338 yGlyProLeuMetGly 343
 seq_name: SwissProt_40:PPSA_PYRAB
 seq_documentation_block:
 ID PPSA_PYRAB STANDARD; PRT; 819 AA.
 AC 09V2H7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water
 DE dikinase) (PEP synthase).
 DE PPSA OR PAB0057.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID:29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ORSAV;
 RA Hellig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
 RT structure and evolution.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -I CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
 CC phosphoenolpyruvate + phosphate.
 CC -I PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
 CC LACTATE ARE USED AS A CARBON SOURCE.
 CC -I SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC
 DR EMBL: AJ248283; CAB49021.1; -
 DR HSP: P22983; IDIK
 DR InterPro: IPR000121; PEP_utilizers.
 DR InterPro: IPR002192; PPK_N_term.
 DR Pfam: PF02896; PEP_utilizers; 1.
 DR Pfam: PF01326; PPK_N_term; 1.
 DR ProDom: PD000940; PEP_utilizers; 1.
 DR PROSITE: PS00370; PEP_utilizers; 1.
 DR PROSITE: PS00742; PEP_ENZYMS; 2; 1.
 DR Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
 FT MOD_RES 441 441 PHOSPHORYLATION (BY SIMILARITY).
 FT DOMAIN 809 816 POLY-GUT
 SO SEQUENCE 819 AA; 90652 MW; E09486AB527C6EA CRC64;

alignment_scores:
 quality: 113.50 length: 455
 Ratio: 0.473 gaps: 21
 Percent Similarity: 52.747 Percent Identity: 21.538

alignment block:

US-09-303-518d-125 x PPSA_PRRAB

Align seg 1/1 to: PPSA_PRRAB from: 1 to: 819

```

67  GCGCCGCGCTTACCGAGTCGCGTGTGGCGAAGATATGCGGAT 116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 GtGluAlaValAlaSerGlyAlaValThrProAspGluTyrIleValAl 268
    : : : : : : : : : : : : : : : : : : : : : :
117 GCGCCGCGCTGAGTAAAGTCAGAGAGCGATCCGTCAGAAAAGGCGAAG 166
    : : : : : : : : : : : : : : : : : : : : : :
268 uLysGlyThrTyrLysIleLysGlu...LysValIleAlaLysLysGluV 284
    : : : : : : : : : : : : : : : : : : : : : :
167 TCCTGTTGAGACAAAGATCCGGCGGTGTTCATGCGCGGCT 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 almet.....ValIleArgAsnProGlu 291
    : : : : : : : : : : : : : : : : : : : : : :
217 TCAGCAAA.....ATCGCGGATTCACCGTGGC..... 246
    : : : : : : : : : : : : : : : : : : : : : :
292 ThrGlyLysGlyThrValThrValLysValAlaGluTyrLeuGlyProG 308
    : : : : : : : : : : : : : : : : : : : : : :
247 .....GAAGCGCGTACTT.....CACTCACTCGTGAATGCGGTTG 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 uTrpValGluLysGluValLeuThrAspGluGluIleIleGluValAla 325
    : : : : : : : : : : : : : : : : : : : : : :
284 AAGGACACACCAATCGATTGTAACCTAGCA...CTGAGCGCTG 330
    : : : : : : : : : : : : : : : : : : : : : :
325 ymerGlyGluIleGlu...GluHisTyrGlyTyrProGluAspIle 340
    : : : : : : : : : : : : : : : : : : : : : :
331 GCA...AACTTAAGCGCGGAGAGAGTGGCGCAACTGATTCGAATCGG 377
    : : : : : : : : : : : : : : : : : : : : : :
341 GluTrpAlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerAr 357
    : : : : : : : : : : : : : : : : : : : : : :
378 TTTGTGACATGCGGTGCGACCGCTCGCTTCAGCAAAATTCGCGCTG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 gProValThrThrLeuLysGluThrThrThrGluGluValGluGluVal 373
    : : : : : : : : : : : : : : : : : : : : : :
428 ATGCGGACCGCTGCGCATCTTCTGCAATGCGATGAGACCAATCG... 474
    : : : : : : : : : : : : : : : : : : : : : :
374 .....GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGly 388
    : : : : : : : : : : : : : : : : : : : : : :
475 CTGGTCCCGACCTACGGTCATATCAAGAGCCCGCAGATTTCAA 524
    : : : : : : : : : : : : : : : : : : : : : :

```

```

389 IleGlyAlaGlyArgValAlaValIlePheAspAlaSerGlu..... 402
525 ACGGCGCGCTGTGGTATTTGACCCGTTGACCGAAGCAAAATTCAGTTT 574
    : : : : : : : : : : : : : : : : : : : : : :
403 .....IleAspLysValLysGluGlyAspValLeuVal 414
    : : : : : : : : : : : : : : : : : : : : : :
575 GTAAGGACGTGGCGCAGAC...GTGCGGCT...GAATGCTGGCAAC 618
    : : : : : : : : : : : : : : : : : : : : : :
414 ThrThrMetThrAsnProAspMetValProAlaMetLysAlaAlaVal 430
    : : : : : : : : : : : : : : : : : : : : : :
619 ATCGAACAACATGATTTGCGCGCGCGCATCTCGCGGTTGAGTGCAC 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 IleIleThrAspGluGlyGlyArgThrSerHisAlaAlaIleValSerAr 447
    : : : : : : : : : : : : : : : : : : : : : :
669 GCACATTCATTATCGACCGCGCGGCGGCAATAAACCGTGGACCA 718
    : : : : : : : : : : : : : : : : : : : : : :
447 gLueGluGlyIleProAlaValAlaGlyThrLysGluAlaThrLysL 464
    : : : : : : : : : : : : : : : : : : : : : :
719 TCAATTAACAAGATGTAATTACCATTTGC..... 747
    : : : : : : : : : : : : : : : : : : : : : :
464 euLysThrGlyAspTyrValThrValAspGlyThrArgGlyLeuVal 480
    : : : : : : : : : : : : : : : : : : : : : :
748 .....CGTTGTTGCAACAGCGCGCTGCAACCGCGCGCT 785
    : : : : : : : : : : : : : : : : : : : : : :
481 LysGlyIleValLysSerLeuValGluLysLysLysGluAlaAla 497
    : : : : : : : : : : : : : : : : : : : : : :
786 GATTGCGCTAGTGTCTTCAAGTCAACAAACCGCGCGCTGCGTACG 835
    : : : : : : : : : : : : : : : : : : : : : :
497 alaAlaProGlyAlaAlaValAlaAlaProLeuValThrGlyThrL 514
    : : : : : : : : : : : : : : : : : : : : : :
836 TTTTGGGTGCAAAAGTATGCAAAATTTACTGCGGCGAAATGGTTGAC 885
    : : : : : : : : : : : : : : : : : : : : : :
514 euValLysValAsnValSerMetProGluValAlaGlu..... 526
    : : : : : : : : : : : : : : : : : : : : : :
886 GACACCGCGGTGATTCGCGTTCGATTTGAACGCGCGGATTCACAAAG 935
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 .....ArgAlaAlaAlaThrGlyAlaAspGlyValGlyLeuAlaG 541
    : : : : : : : : : : : : : : : : : : : : : :
936 CGCGCAGCTTATTTGGACGCTTACCAATGATTCGTTATTCGAAG 985
    : : : : : : : : : : : : : : : : : : : : : :
541 agLueHisMetIleLeuSerIleGlyGlnHisProValLysPheLeu 558
    : : : : : : : : : : : : : : : : : : : : : :
986 AAGCGCGCGCAAGAGGTGCGGTGCGGTGGCGCGCGCGCGACGAA 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
558 LuGlyLysGluGluGluLeuValGluLysLeuAlaGluGlyIleGlu 574
    : : : : : : : : : : : : : : : : : : : : : :
1036 TACTCCATTCACGCTACACCGCTGCGCATTTCTGAAAACAACACTT 1085
    : : : : : : : : : : : : : : : : : : : : : :
575 ValAlaAlaAla..AlaPheTyrProAlaProVal...TyrTyrArgThr 589
    : : : : : : : : : : : : : : : : : : : : : :
1086 CAAGTTTCAACACAGCGCTCAACGCGCGCGCGCGCGATGCGGATG 1135
    : : : : : : : : : : : : : : : : : : : : : :
590 AspAlaProThrAsnGluProPheArgGluMetPro...GlyGlyLys 605
    : : : : : : : : : : : : : : : : : : : : : :
1136 GTACTTACAGCGCGCTGATGCGCTTGATATCTGCG..... 1172
    : : : : : : : : : : : : : : : : : : : : : :
605 uProGluGluArgAsnProMetLeuGlyTyrProGlyIleArgArgGly 622
    : : : : : : : : : : : : : : : : : : : : : :
1173 .....CACCTGCTTTTTCGCGATTTATGTCGCGGATTCGCAAGCG 1217
    : : : : : : : : : : : : : : : : : : : : : :
622 euAspGluProGluLeuLeuArgAlaGluPheLysAlaIleLysVal 638
    : : : : : : : : : : : : : : : : : : : : : :
1218 GCAGCATTTGCTGTTGTAATTTGAGCAAGAAAGACCTGCTTTGCA 1267
    : : : : : : : : : : : : : : : : : : : : : :
639 ValGluLysGlyTyrAsnAsnIleGly.....ValMet 649
    : : : : : : : : : : : : : : : : : : : : : :
1268 GCTTCTGCTG.....CCGCGCAATATGCAATACGCGCGCTGTTG 1308
    : : : : : : : : : : : : : : : : : : : : : :
649 cLeuProLeuValSerHisProGluGlnIleArgLysAlaLysGluIle 666
    : : : : : : : : : : : : : : : : : : : : : :
1309 CGCAAGTGTGCA 1322
    : : : : : : : : : : : : : : : : : : : : : :
666 laArgSerValGly 670
    : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: SwissProt_40:VL2_HPV70

seq_documentation_block:

ID VL2_HPV70 STANDARD; PRT; 466 AA.

AC P50801;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Minor capsid protein L2.

GN L2.

OS Human papillomavirus type 70.

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

OC Papillomavirus.

OX NCBI_TaxID=39457;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96249586; PubMed=8815087;

RA Forslund O., Hansson B.G.; 70 genome cloned from overlapping PCR

RT products: complete nucleotide sequence and genomic organization.";

RL J. Clin. Microbiol. 34:802-809(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U21941; AAC54856.1; -
 DR InterPro: IPR000784; Late_L2.
 DR Pfam: PF00513; Late_protein_L2; 1.
 KW Coat protein; Late protein.
 SQ SEQUENCE 466 AA; 49904 MW; 91E15579035F2377 CRC64;

alignment_scores:

Quality: 112.50 Length: 402
 Ratio: 0.349 Gaps: 21
 Percent Similarity: 50.995 Percent Identity: 24.876

alignment_block:

US-09-303-518D-125 x VL2_HPV70 ..

Align seg 1/1 to: VL2_HPV70 from: 1 to: 466

28 CTGCCCATCGCGGCGAGACCGGACGCGTTTACGAGCGCCGCGCAT 77
 ::::::::::::::::::::: |||
 72 ILeProLeuGlyGlyArgProSerThrValValAspValThrProAlaAr 88
 78 TACCGAAGTCGCGTCTGGGGAAGATATGCGCGTATGCGGCCCTGGA 127
 ::::::::::::::::::::: |||
 88 gProProValValIle.....GluProValJelyProThrgluProSerI 103
 128 TG.....AAGTCAAGAGAAGCGATGCGTCAAAAAAGGCCAAGTGTG 171
 ::::::::::::::::::::: |||
 103 IeValGlnIleuValGlnGluSerSerValValSerSercly.....I 116
 172 TTTGAAGACAAAAAGATCCGGGCGTGTGTACTGCGCGCGCTTCAGG 221
 ::::::::::::::::::::: |||
 117ThrProIleProThrPheThrGlyThrSerclyph 128
 222 CAAAAATCGCGCGATTCACCGTGGGGAAGCGGACTTACGTCAGTCG 271
 ::::::::::::::::::::: |||
 128 eguIleThrSerSerAlaThrThrThrProAlaValIleuAsp.....I 143
 272 TGAATCCGTTGAGGCAACGAGAAATGAGTTGACGCTACGCACT 321
 ::::::::::::::::::::: |||
 143 IeThrProAlaSerGlySerValGlnIleSerThrThrSerclyThrAsn 159
 322 GAAGCGCTGGCAAACTTAAAGCGGCGAAGATGCGCGCAACCTGATCA 371

160 ProAlaPheAlaAspProSerLeuIleGluValProGlnThrGlyGluVal 176
 ::::::::::::::::::::: |||
 372 ATCCGGTTTGTGACGTGCGCGACCGCGCGCTGACGAAATTCCTG 421
 ::::::::::::::::::::: |||
 176 IserGly.....AsnIlePheValThrThrProThrSerclyThrIsc 191
 422 CCGTCGATCGCGAGCGCGTTCGCCATCTCTGTC.....AATGCGATGACACC 468
 ::::::::::::::::::::: |||
 191 IYTYGlnGlnIleProMetGlnValPheAlaSerHisGlyThrGlyThr 207
 469 AATCGCGTGGCTGCGCGACCTACGCTCATTTATCAAAAGACCGCGGAGA 518
 ::::::::::::::::::::: |||
 208 GluProIleSerSerThrProValProGlyValSerAlaGlyVal.. 223
 519 TTTCAAGCGCGCGCTGTGATTATGACCGCTTGACGCAACCAAAATCC 568
 ::::::::::::::::::::: |||
 224ProArgLeuTySerArgAlaTyH 232
 569 ATGTTTGTAAAGCAGCTGCGCGACGAGCTGCGCTGAAATGCTGCCAAC 618
 ::::::::::::::::::::: |||
 232 IeGlnValArgValAsnAspPheAspPheValThrArgProSerSerPhe 248
 619 ATCGAAACACATGATTCGGCGCGCG.....CATCGCGCGTGTAG 662
 ::::::::::::::::::::: |||
 249 Val.....ThrPheAspAsnProAlaPheGluProGlyAspThrSe 262
 663 TGGCAGCGCATTTTCATTCAGACCGCGCGCGCAATTAACCGCTGT 712
 ::::::::::::::::::::: |||
 262 rIeuthr.....PheGluProAla.....AspThrAlaP 272
 713 GGCACATCAATTATCAAGATGTAATTACATTCGCGGTTGTTGTCACAA 762
 ::::::::::::::::::::: |||
 272 roAspProAspPheLeuAspIleValArgLeuHisArgProAlaLeuthr 288
 763 GCGCGCTGACACACGCGCGCGTGAATTCGCTAGGTGTCTCAAGTCAA 812
 ::::::::::::::::::::: |||
 289 SerArgArgGlyThrValArgPheSerArgLeuGlyLysAlaIleThr 305
 813 CAACCGCGCTCTGCTGACCGTTTGGTGGCGAAAGTA..... 852
 ::::::::::::::::::::: |||
 305 tPheThrArg...ArgGlyThrGlnIleGlyAlaGlnValHisTyTYH 321
 853TGCCAATTACTGCGCGCGCAATTCGCTGACACACACCGC 894
 ::::::::::::::::::::: |||
 321 IsAspIleSerAsnIleThrAlaThrGluAspIleGluMetGlnProLeu 337
 895 GTGATTTCCGCTTCGCTATTTGAACGGCGGATTCACAAAGCGCGCAGA 944
 ::::::::::::::::::::: |||
 338 LeuThrSerLeuSer.....ThrAspGlyLeuTyHAs 348
 945 TTAATTGG...GACGCTACCAATCAGATTTCGCTATTCGAAAGAGGCG 991
 ::::::::::::::::::::: |||
 348 P.IleTyAlaAspAlaAspIleAspAsnAlaMetLeuHisThrThrSe 364
 992 GCAGCAAGAGCGTTCGCGCTGGTTCGCGCGCGGCAACCAATCTCC 1041
 ::::::::::::::::::::: |||
 365 HisThrGlySerThrGlyPro.....ArgSerHisLeuSerPhePr 378
 378 oSer.....IleProSerThrValSer.....ThyLysTySerAsn 391
 1092 CAACACAGCGCTCAAGCGCGCG.....ACCGGCGCA 1123
 ::::::::::::::::::::: |||
 391 hrThrIleProPheThrThrSerThrAspIleProValThrThrGlyPro 407
 1124 TGTGCGGATTCGCTTACGAGCGCGTGAATGCCCTTGATATCCGCC 1173
 ::::::::::::::::::::: |||
 408 AspIleValIleuProThrAlaSerProAsnLeuProPheVal.....Pr 422
 1174 ACCC 1177
 |||


```

236 ethnglymetlylsasnThrspillethrgluThrHislysmetArgThr 253
836 TTTTG.....GTCGCAAAATATCGCAAT... 861
253 lueulysasphsAlaAlaAlaSerGlyllelyleuThrYrleAla 269
862 .....ACTGGGGCAATTGGTTGACACAGACAGACCGCGT 896
270 PheilellelysalValAlaLysSerleuArgAspMetProAsnIleas 286
897 GATTTCGGTTCGTAATGACGGCCGATT..... 927
286 valArgGlyAspPheAlaAsnAsnLysIleGlnPheMetHisAsnIleA 303
928 .....ACA 930
303 snlleGlyIleAlaValAspThrProAsnGlyLeuMetValProValIle 319
931 CAAGGGCGGACGATTTATTTGGACGCTACACAAATCAGATTTCGCTAT 980
320 LysGlyAla...AspHisLeuSerValPheGluIleAlaIleLysIleSe 335
981 CGAAGAAGCGCGGACGAAGAGCTGTTCGGCTGCGCGCGGACCGCG 1030
335 rglu.....LeuAlaAsnLysAlaL 342
1031 ACAAAATATCCATCAGCGCTACA.....ACCGTC 1059
342 ysAspGlyLysleuThrArgAlaGluMetThrGluAlaIleThrPheThrVal 358
1060 GGCGATTTCTGTAACAAACAACTCTTCAAGTTCAACACAGCGCTACG 1109
359 SerAspPheGlySerValGlyLeuAspYrAlaIleThrProIleIleAsnSe 375
1110 CGCGACCGCGCCATGTCGCTGCTACTTACGAGCG..... 1149
375 rProGluSerAlaIleLeuGlyValGlyThrMetSerGlnThrProLeuT 392
1150 .....GTGATCCCTTGGATATC 1167
392 YrIleAsnGlyLeuGluGlnLysArgPheIleMetProLeuSerMet 407
seq_name: SwissProt_40: FASL_SCHPO
seq_documentation_block:
ID FASL_SCHPO STANDARD; PRT; 2073 AA.
AC Q90UG0; P78799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid synthase subunit beta (EC 2.3.1.86) [Includes: 3-
DE hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);
DE Enoyl-[acyl-carrier-protein] reductase (EC 1.3.1.9); [ACYL-
DE carrier-protein] acetyltransferase (EC 2.3.1.38); [ACYL-carrier-
DE protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase
DE thioesterase (EC 3.1.2.14)].
DE FASL OR SPAC926.09c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SUBUNIT.
RX MEDLINE=98359895; PubMed=9593066;
RA Niwa H., Katayama E., Yanagida M., Morikawa K.;
RT Cloning of the fatty acid synthetase beta subunit from fission yeast,
RT coexpression with the alpha subunit, and purification of the intact
RT multifunctional enzyme complex";
RL Protein Expr. Purif. 13:403-413(1998).
RP SEQUENCE FROM N.A.
RP STRAIN=972;

```

```

RA Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1725-2073 FROM N.A.
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioke S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RT CDNA's.";
RL DNA Res. 4:363-369(1997).
CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN]
CC ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID
CC SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND
CC 3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.
CC CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =
CC -1- LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-
CC 2,3-dehydroacyl-[acyl-carrier protein] + NADH.
CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC carrier protein] + oleate.
CC -1- SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(5)] HEXAMERS OF
CC TWO MULTIFUNCTIONAL SUBUNITS (ALPHA AND BETA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AB010274; BA36384.1; -
DR EMBL: AB110469; CAB54157.1; -
DR EMBL: D89148; BA13810.1; -
DR InterPro: IPR001227; Acyltransf domain.
DR InterPro: IPR003965; Fatty acid synth.
DR InterPro: IPR002539; Maoc dehydratase.
DR Pfam: PR00698; Acyl_transf. 1.
DR Pfam: PR01575; Maoc-dehydratase. 1.
DR PRINTS: PR01483; FASYNTHASE.
KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;
KW Transferrase; Hydrolase; Lyase; NAD; NADP.
FT DOMAIN 1 459 ACETYL TRANSFERASE.
FT DOMAIN 470 858 ENOYL REDUCTASE.
FT DOMAIN 1155 1644 DEHYDRATASE.
FT DOMAIN 1645 2073 MALONYL/PALMITOYL TRANSFERASE.
FT ACT_SITE 270 270 ACETYL TRANSFERASE (BY SIMILARITY).
FT ACT_SITE 1361 1361 DEHYDRATASE (POTENTIAL).
FT ACT_SITE 1828 1828 MALONYLTRANSFERASE (BY SIMILARITY).
FT CONFLICT 222 P -> R (IN REF. 1).
SQ SEQUENCE 2073 AA; 23055 MW; D92827083BE7C7C4 CRC64;

```

alignment_scores:

Quality: 110.50 Length: 465
Ratio: 0.521 Gaps: 25
Percent Similarity: 45.591 Percent Identity: 20.215

alignment_block:

US-09-303-518d-125 x FASL_SCHPO
Align seg 1/1 to: FASL_SCHPO from: 1 to: 2073

```

217 TCAGCGAAATATCGCCGCGATTCACGCTGCGCAAAACGCGCTACTTCAGTC 266
:::|||||::: :::::||||| :::::|||||

```

```

948 ThrGlyLysIeu.....MetArgArgIleGluGluArgPheThrGlnAs 962
267 AGTC.....GTGATTGCCGTTGAGAGCAGCAGCAATGAG..... 303
    |||      |||      |||      |||      |||      |||
962 pValGlyLysThrThrLeuIleuAsnPheGluAspLeuAsnAspProt 979
304 .....TTTGACGCTACGACGACCTGAGCGCTGGCA 333
    |||      |||      |||      |||      |||      |||
979 yrrProValAlaAlaArgPheLeuAsnPalatyrProGluAlaSerThrGln 995
334 AACTTAAGCGGCGAAGAGTGGCCGCAACCTGATCCAAATCCGGTTGTG 383
    |||      |||      |||      |||      |||      |||
996 AspLeuAsnThrGlnAspAlaGln.....PhePh 1005
384 GACGCGCTGGCGCGCGCTCGCTGACGAA.....ATTGCTG 421
    |||      |||      |||      |||      |||      |||
1005 eTyrSerLeuCysSerAsnProPheGlnIysProValProPheIleProA 1022
422 CCGTCGAT..... 429
    |||      |||      |||      |||      |||      |||
1022 LaileAspAspThrPheGluPheThrPheLysAspSerLeuTrpGln 1038
430 GCCGAGCCGCTGCCATCTTGTGTAAT.....GGAT 461
    |||      |||      |||      |||      |||      |||
1039 SerGluAspLeuAlaAlaValAlaGlyGluAspValGlyArgValAlaAl 1055
462 GGACACCAATCCGCTGGCTGCCGACCTACGATTCATTCAGAACAGCCG 511
    |||      |||      |||      |||      |||      |||
1055 eleuGlnIysProMetAlaAlaIleLysHisSerThrLysValAsnIleProA 1072
512 CCGAGGATTTCAACCGCGCGCTGTGTATAGACCCGTTGACGAGCAAGC 561
    |||      |||      |||      |||      |||      |||
1072 LaLysGlu.....LeuLeuAspGlyIleAsnGluThr 1082
562 AAAATC...CATGTTTGTAGGCA.....GCTGGCCGAGAC.....GT 596
    |||      |||      |||      |||      |||      |||
1083 HisIleGlnHisPheIleLysLysPheThrLysIleAspGluLysIle 1099
597 GCCGCTGAAATGCTGCCACATCGAACAATTCGGCGCGCGC 646
    |||      |||      |||      |||      |||      |||
1099 eProIle.....ValGluTyrPheLysIleValP 1109
647 ATCTGCGCGTTTGTAGT..... 663
    |||      |||      |||      |||      |||      |||
1109 rProValAlaHisValSerHisLysSerLeuGlnSerValThrGlu 1125
664 .....GGCAGCGACATTCATTCATTCAGCGCGCGCGGAGTAAC 707
    |||      |||      |||      |||      |||      |||
1126 GluAlaGlySerLysValTyrLysLeuProGluIleGlySerAsnSerAl 1142
708 CGTG..... 711
    |||      |||      |||      |||      |||      |||
1142 AleuProSerLysLysLeuTyrPheGluLeuLeuAlaGlyProGluTyrT 1159
712 .....TGAC 717
    |||      |||      |||      |||      |||      |||
1159 hTrpPheArgAlaIlePheThrThrGlnArgValAlaLysGlyTyrLys 1175
718 ATCAATTTATACAGATGTATATACATTCGCGCTTGTTCAGACGCGCG 767
    |||      |||      |||      |||      |||      |||
1176 LeuGluHisAsnProVal.....ArgArgIlePheAlaPro..... 1187
768 TCTGAACACCGACGCGCTGATTCGCTTGTGTTCTCAAGTCAACAAAC 817
    |||      |||      |||      |||      |||      |||
1188 ArgTyrGlyGlnArgAlaValValLysGlyLysAspAsnPalatyrVal 1204
818 CCGCGCTCTTGGGTCAC.....GTTTGGGGTCCGAAAGTA 852
    |||      |||      |||      |||      |||      |||
1204 alGluLeuTyrGluThrGlnSerGlyAsnTyrValLeuAlaAlaArgLeu 1220
853 TGGCAAAATTACGCGGCGCAATTGTTGTCACACAGACCGGTCATTTC 902
    |||      |||      |||      |||      |||      |||
1221 Ser.....TyrAspGlyGluThrIleValValSe 1230

```

```

903 CGGTTGCTATTGAACGGCGGATTCACAGCGCGCAC...GATTATT 949
    |||      |||      |||      |||      |||      |||
1230 rMetPheGluAsnArgAsnAlaLeuLysGlyGluValHisLeuAspPheI 1247
950 TGGGACGCTAC.....CACATTCAGATTTCGGTTATCGAA 984
    |||      |||      |||      |||      |||      |||
1247 euPheLysTyrGluProSerAlaGlyTyrSerProValSerGluIleLeu 1263
985 GAAGCGCGGACGCAAGACGTG.....TTGCGCTGGCTT..... 1017
    |||      |||      |||      |||      |||      |||
1264 AspGlyArgAsnAspArgIleLysHisPheThrThrAlaLeuTrpPheG 1280
1018 ....GGCGCGACGCGGACAAATTCATTCACGCGTTCACAACTCCGCC 1063
    |||      |||      |||      |||      |||      |||
1280 yGluGluProTyrProGluAsnAlaSerIleThrAspThrPheThrGlyP 1297
1064 ATTTCTGTAAA.....AACAAAGCTTCATTCACACACAGCCGTC 1104
    |||      |||      |||      |||      |||      |||
1297 roGluValThrValThrGlyAsnMetIleGluAspPheCysArgThrVal 1313
1105 AACGCGCGGACGCGCGCATGTGCGGATTTGCTACTTACGACGCGGTAT 1154
    |||      |||      |||      |||      |||      |||
1314 GlyAsnHisAsnGluAlaLysThrLysArgAlaIleArgLysArgMetAl 1330
1155 GCCCTTGGAT.....A 1165
    |||      |||      |||      |||      |||      |||
1330 aProMetAspPheAlaIleValAlaGlyTyrGlnAlaIleThrLysAlaI 1347
1166 TCGTGGCCGACCGCTGTTTGGCGGATTTA.....ATCGTCGCGCATTC 1209
    |||      |||      |||      |||      |||      |||
1347 LeuPheProLysAlaIleAspGlyAspLeuLeuArgLeuValHisLeuSer 1363
1210 GACAGCGCGGACGATTTGGTCTTGTGAATTCAGCAGAGAGAGAC 1254
    |||      |||      |||      |||      |||      |||
1364 AsnSerPheArgMetValAlaGlySerHisSerLeuMetGluGlyAsp 1378
seq_name: SwissProt_40:CHLN_SYNP7
seq_documentation_block:
ID CHLN_SYNP7 STANDARD; PRT; 466 AA.
AC P54208;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit N
DE (EC 1.18.-.-) (Lr-POR subunit N) (DPOC subunit N).
DE CHLN OR FRXC.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95388763; PubMed=7659748;
RX Ronen-Tarazi M., Lieman-Hurwitz J., Gabay C., Orus M.I., Kaplan A.;
RT "The genomic region of rbcLs in Synechococcus sp. PCC 7942 contains
RT genes involved in the ability to grow under low CO2 concentration and
RT in chlorophyll biosynthesis."
RL Plant Physiol. 108:1461-1469(1995).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC protochlorophyllide (Pchl) to form chlorophyllide a (Chlide)
CC (By similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits: chlL, chlN and chlB. Could form a
CC heterotrimer of two chlB and two chlN subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE BCCH / CHLN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X67694; CAA47924.1;
KM Photosynthesis: Chlorophyll biosynthesis; Oxidoreductase.
SEQUENCE 466 AA; 51540 MW; 981A51305045574C CRC64;

alignment_scores:

Quality: 107.00 Length: 333
Ratio: 0.728 Gaps: 16
Percent Similarity: 44.144 Percent Identity: 21.021

alignment_block:

US-09-303-518d-125 x CHLN_SYNP7

Align seg 1/1 to: CHLN_SYNP7 from: 1 to: 466

```

7  AAAAAAAAAAGTTAACTGCGCCATCGCGGCGAGACCGGACGAGC 56
   |||
92  GlnlleYsArGaSPaRgaSn.....ProSerValIl 102
   |||
57  CGTTACGACGCGCGCCATTCAGCAAGTCGCTTGGTGGCAAGAAAT 106
   |||
102  evalTrpIleGlyThrcysThrThGlnlleYlsYsMetAspLeuGln 118
   |||
107  ATGCCGGTATGCGCCCTTCGATGAAGTCAGGAGGCGATCCGTCMAA 156
   |||
119  ....GlyLeuAlaProIysLeuGlnIleGlyIlePro..... 131
   |||
157  AAGGCGAAGTGCTGTTGAAGACAAAAGAAATCCGGCGGTGGTGTAC 206
   |||
132  .....IleValva 134
   |||
207  TCGCGCGCTTCAGCAAAATCGCGCATTCACCGTGCAGAAAGCGCG 256
   |||
134  lAlaTrGlnAlaSnGlyLeuAspTYrAlaPheThGlnIYgIlnsPhrY 151
   |||
257  TACTTCAGTCACTCGTGAATCCCGTTGAAGCAAGCAAGAAATGACTTT 306
   |||
151  AlLeuAlaIleMetAlaIaArg..... 158
   |||
307  GAACGCTACGACACGCGCTGAGCGCTGAACCTTAAGCGCGGACGAGTGC 356
   |||
159  .....CysProGlnAlaIaIaThrSerGlnAlaAspGlnGlnuar 172
   |||
357  CCGCAACCTGATCCCAATCCGGTTGTGACTGCGCTCGACCCGCTCCGT 406
   |||
172  gThrAsnAlaIleGln.....ArgLeuLeuGlnP 182
   |||
407  TCAGCAAAATTCCTCCGCTGATCCGAG..... 435
   |||
182  heGlyLysSerProAlaIaGlnGlnInProAlaSerSerIYsHISPro 198
   |||
436  CCGTTCGCGATCTTCGTCATGATGAGACCAACCAATCCGCTGCTCCCA 485
   |||
199  ProLeuIleLeuPheGlySerValProAsp.....ProAlaIaThrG 213
   |||
486  CCTTACGCTCATTTATCAAGAAGCGCGGATTTCAACGCGGCTGT 535
   |||
213  nLeuThrIle.....GlnLeuAlaIYsGlnGlyIleT 224
   |||
536  TGTATTGAGCGCTTGACGCAAGCAAAATCCATTTGTGAAGCAGCT 585
   |||
224  hrValSerGlyTrpLeuProAlaIYsArgYr..... 234
   |||
586  GCGCGACAGCTGCCG.....TGTGAATAATGCTGCCACATGCAAAACAC 629
   |||
235  ...ThrcIleuProValIleAlaIaGlnGlySerTYrAlaIleGlyLeu 250
   |||
630  TGAATTC..... 636
   |||
250  nProPheLeuSerArgThrAlaIaThrThrLeuMetArgArgIYsCysL 267

```

seq_name: SwissProt_40:AMYH_YEAST

seq_documentation_block:

ID AMYH_YEAST STANDARD; PRT; 1367 AA.

AC P08640; P08068;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)

GN STAI OR STAZ OR MAL5 OR YIR019C.

OS Saccharomyces cerevisiae (Baker's yeast)

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN NCBI_TaxID=4932;

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,

RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,

RA Gentles S., Hamlyn N., Hornsby T.S., Hunt S., Jagsels K., Jones M.,

RA Louis E., Lye G., Mould S., Mould T., Odeh C., Pearson D.,

RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,

RA Walsh S.V., Whitehead S.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

RX MEDLINE=87194600; PubMed=3106330;

RA Yamashita I., Nakamura M., Fukui S.;

RT "Gene fusion is a possible mechanism underlying the evolution of

STAI.;"

RL J. Bacteriol. 169:2142-2149(1987).

RN [3]

RP SEQUENCE OF 1-31 FROM N.A.

RC STRAIN=SPX101-1C;

RX MEDLINE=89031230; PubMed=3141213;

RA Pardo J.M., Ianez E., Zalacain M., Clares M.G., Jimenez A.;

RT "Similar short elements in the 5' regions of the STAZ and SGA genes

from Saccharomyces cerevisiae.;"

FEBS Lett. 239:179-184(1988).

CC -1- CARBOLYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

glucose residues successively from non-reducing ends of the chains

with release of beta-D-glucose.

CC -1- SIMILARITY: TO S.POMBE SPBC215.13

CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL; Z38061; CA86176.1; -
 DR EMBL; M16164; AAA35014.1; -
 DR EMBL; M16165; AAA35015.1; -
 DR EMBL; X13857; CAA32069.1; -
 DR PIR; B26877; B26877.
 DR PIR; A26877; A26877.
 DR PIR; S48478; S48478.
 DR SGD; S0001458; M0C1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1367 GLUCOMYLASE SL/SZ.
 FT DOMAIN 210 1367 SER/THR-RICH.
 FT CARBOHYD 817 817 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAc. . .) (POTENTIAL).
 SQ SEQUENCE 1367 AA; 136110 MW; 91C0E2DBD61AA9D CRC64;

alignment_scores:
 Quality: 107.00 Length: 545
 Ratio: 0.476 Gaps: 18
 Percent Similarity: 41.284 Percent Identity: 19.633

alignment_block:

US-09-303-518D-125/rev x AMYH_YEAST ..

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

1337 TCCTTCAATGTTTCCAGCACTTTCGCAACAGCGGCGGATTCGTA 1288
 |||||
 891 SerheserThrGlyThrValThrProserSerSerlySTY.... 905
 1287 TTTGCCGGGAGACGAGCGCAAGAGAGTCTTCGTCGCAAT 1238
 |||||
 906ProGlySerInThrGluThrSerSerThrThrGluThr 921
 1237 CCAAGCAACCAATGCTGCGCTGCGTA.....TCGCG 1200
 |||||
 921 hrleValProThrLysThrThrSerValThrThrProserThrThr 937
 1199 ACGATTAATCGCGCAAGAGGAGGATATCCAAAGGCAATC 1150
 |||||
 938 ThrThrThrThrValCysSerThrValThrSerSerValGluThr 954
 |||||
 1149 GCGCTCGTAAGTACCAATCGGCGGCGGCGGCGGCGGCGG 1100
 |||||
 954 rThrSerGlySerProLysThrValThrThrValProthrThr 971
 1099 CTGCTGTAAGTGAAGATTGTTTTCAGAAATGCGCGAGGTTGTA 1050
 |||||
 971 hrThrThrSerValThrThr.....SerSerThrThrThrThr 983
 1049 CCGCGATGAGGATTTTCGCGGCGGCGGCGGCGGCGGCGG 1000
 |||||
 984 ThrThrThrValCysSerThrGlyThrThrSerValGluThrThrSe 1000
 |||||
 999 TTTGCTGCGGCGCTTCGATACGGAATCGATTGTAGCGTCCA 950
 |||||
 1000 rGlySerProLysThrThr..... 1008
 949 AATAATCGGCGGCTGTGTAATCGCGGCTTCATACGGAACGGA 900
 |||||
 1009ThrThrValProCysSerThrSerProserGluThrAlaSerGlu 1023
 |||||
 900 900
 1024 SerThrThrThrSerProThrThrProValThrThrValThrThr 1040

899ATCAGCGGCTGTCTGTCACCAACCAATTCGCCGAGTAATT 857
 |||||
 1040 rValValThrThrGlyThrSerThrSerThrLysProGlyGluThr 1057
 |||||
 856 GCGATCTTTCGACCAACCAAGGTAACGAGCGCGGTTGTGACT 807
 |||||
 1057 hrThrThrThrThrThrThrLysAsnThr....ProThrThrThrThrThr 1072
 |||||
 806 TGAGAACCACTTAGGCAATCAGCGCGTGTTCAGAGCGCTGTTC 757
 |||||
 1073 lLeaLaproThrProSerValThr.....ThrValThr 1083
 |||||
 756 AAACAACGCGCAATGTAATTCATCTGATATGATGATGTCACACG 707
 |||||
 1083 rAsnPhenThrProThrThrThrThrThr.....ThrV 1094
 |||||
 1094 aLysSerThrGlyThrAsnSer.....AlaGlyGluThr 1105
 |||||
 656 CCGGAGGATGCGGCGG..... 639
 |||||
 1106 ThrSerGlySerProLysThrValThrThrValProCysSerThr 1122
 |||||
 639 639
 1122 rGlyThrGlyGluThrThrThrGluAlaThrThrLeuValThrThrAlaV 1139
 |||||
 639 639
 1139 aThrThrThrValThrThrThrGluSerThrThrGlyThrAsnSerAla 1155
 |||||
 638CCGAATTCATGTGT 625
 |||||
 1156 GlyLysThrThrThrGlyThrThrThrLysSerValProThrThrThrYVal 1172
 |||||
 624 TTCGATGTTGCGACATTTTCAGACGCGACGTCGCGCAGCTGCTAC 575
 |||||
 1172 lThrThrLeuAlaProSerAlaProValThrProAlaThrAsnAlaValP 1189
 |||||
 574 AAACATGATTTTGGCTGCGTCAACAGGCTCAATACCAACAGCGCGGT 525
 |||||
 1189 roThrThrThrThrThrThrGlyCysSerAlaAlaThrAsnAlaThrAlaGly 1205
 |||||
 524 TTGAATCCTCGCGGCGCTTTCATATGACCGTAGGCTGCGGCGCAG 475
 |||||
 1206 GluThrThrSerValCysSerAlaLysThrThrLeuThrSerThrAlaSerAla 1222
 |||||
 474 CGGATTTGGTCCATCGCATTCGACGAATGCGCAACGCGCTGCGCATGA 425
 |||||
 1222 acGly.....GluAsnThrAlaProS 1229
 |||||
 424 CCGCAGGAATTTTGTCAACGAGCGGTCGCGACGCGACGTCACAAACCG 375
 |||||
 1229 erAlaThrThrProValThrThrAlaThrProThrThrVal..... 1242
 |||||
 374 GATGATGATGATGCGCGCGCACTTTCGCGCGCTTAATTTGCCAGCGC 325
 |||||
 1243lThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1254
 |||||
 324 TTCAGTGGGTAGCTTCGAACGCGGTCGTCGCTTCGCTTCACGCGCA 275
 |||||
 1254 agGlyGluThrThrThrGlyThrThrThrThrThrThrThrThrThrThrThr 1271
 |||||
 1271 lThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1274
 |||||
 224 TTGCTT.....GAAGCGGCGCGGTAACAC 199
 |||||
 1275 lLeProGlySerAsnGlyAlaLysAsnThrGluThrThrValAlaThrAlaThr 1291

```

198 CACGCC.....GGATCTTTTGTCTT 176
1291 rnsnProIleSerIleThrThrSerGlnLeuAlaThrAlaSerA 1308
175 CAACAGCACTGGCCTTTTGGAGCATGGCTTCTTCTTCTTCTTCTT 126
1308 laseSerValAlaProValAlaThr...SerProSerLeuThrGly... 1322
125 GAGGGCGCATATTCCTTCGCCAAGCAAGCAAGCAAGCAAGCAAG 76
1323 .....ProLeuGlnSerIleAlaSerGlySerAlaValAlaThrTy 1335
75 GGCCGGCCGCTGTAAACGGCT.....TGCTCGGCTGTGCGCG 38
1335 rSerValProSerIleSerSerThrTyrGlnGlyAlaAlaAlaSerIle 1352
37 CGATGGCAGGTTTAGACCTTTTGTGATTTAATC 3
1352 allLeuGlyAsnPhenMetTrpLeuLeuAlaLeu 1363

```

seq_name: SwissProt_40:PPSA_PYRFU

seq_documentation_block:
ID PPSA_PYRFU STANDARD; PRT; 817 AA.

AC P42850; Q59672;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water dikinase) (PEP synthase).
PSA OR PF0043.
GN Pyrococcus furiosus.
OS Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
NCBI_TaxID=2261;
OX RN
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Vc1 / DMS 3638;
RX MEDLINE=95129854; PubMed=7828869;
RA Robinson K.A., Schreier H.J.;
RT "Isolation, sequence and characterization of the maltose-regulated mltA gene from the hyperthermophilic archaeum Pyrococcus furiosus.";
RL Gene 151:173-176(1994).
[2]
RN RP
RP SEQUENCE FROM N.A.
RX STRAIN=Vc1 / DMS 3638;
RX MEDLINE=95354939; PubMed=7628701;
RA Jones C.B., Fleming T.M., Piper P.W., Littlechild J.A., Cowan D.A.;
RT "Cloning and sequencing of a gene from the archaeon Pyrococcus furiosus with high homology to a gene encoding phosphoenolpyruvate synthetase from Escherichia coli.";
RL Gene 160:101-103(1995).

CC CC
CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC phosphoenolpyruvate + phosphate.
CC -1- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC LACTATE ARE USED AS A CARBON SOURCE.
CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U08376; AAA81512.1; -
CC EMBL: X80819; CA56785.1; -
CC HSSP: P22983; IDIK.
CC InterPro: IPR000121; PEP_utilizers.
CC InterPro: IPR002192; PPDK_N_term.
CC Pfam: PF00391; PEP_utilizers_1.
CC Pfam: PF02896; PEP_utilizers_C_1.
CC Pfam: PF01326; PPDK_N_term; 1.

DR ProDom: PD000940; PEP_utilizers; 1.
DR PROSITE: PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE: PS00742; PEP_ENZYMES_2; 1
KW Transferase; Kinase; ATP-binding; Phosphorylation.
FT MOD_RES 442 442
FT DOMAIN 809 815 POLY-GLU.
FT CONFLICT 747 747
SO SEQUENCE 817 AA; 90485 MW; 582694CEFF13C74BA CRC64;

alignment_scores:
Quality: 105.50 Length: 468
Ratio: 0.451 Gaps: 22
Percent Similarity: 50.000 Percent Identity: 21.368

alignment_block:
US-09-303-518D-125 x PPSA_PYRFU ..

Align seg 1/1 to: PPSA_PYRFU from: 1 to: 817

```

67 GGCCCGCCATTCACGAGTCGCTTGGCGAAGATATGCCGTTAT 116
117 GGCCCGCCATTCACGAGTCGCTTGGCGAAGATATGCCGTTAT 166
252 GllGluAlaValAlaSerGlyAlaValAlaThrProAspGluThrIleValG1 268
117 GGCCCGCCATTCACGAGTCGCTTGGCGAAGATATGCCGTTAT 166
268 ulysGlyThrTrpIleLysGluValIleAlaLysGluValM 285
167 TGCTGTTTGAAGACAAAGATCCG.....GGCTGTGTTT 204
285 etValIle.....ArgAsnProGluThrGlyArgGlyThrValMet 298
205 ACTGCCCGGCTTCAGGCAAAATGCCGCGATTCACCGTGC..... 246
299 Val.....LysValAlaGlu...TyrLeuGlyProGluTr 309
247 .....GAAAGCGGCTACTT.....CAGTCAGTCGATTCGCGTTGAG 286
309 PValGluLysGlnValLeuThrAspGluIleIleGluValAlaLysM 326
287 GCAACGACGAATCGATTGAACGCTACGCACTGACCGTACGCGCA... 333
326 etGlyIleLysIleGluAspHisTyrGlyTyrProGlnAspIleGluTrp 342
334 AACTTAACGCGGCAAGATGCGCCGCAACCTGATCAATCCGGTTTGTG 383
343 AlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgProI1 359
384 GACTGGCGTGGCAGCCGCGCTTCAGCAAAATTCCTGCGCGTGCAGTCCG 433
359 ethrThrLeuLysGluGluAlaThrAlaGlu...GluAlaGluGluValG 375
434 AGCGTTCGCGCATTCGTCATGCGATGCGATGACCAATCCG...CTGGCT 480
375 IuGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIleGly 391
481 GCGGACCCCTACGCTCATTTCAAGAGCCCGCGAG.....GATTCAA 524
392 AlaGlyArgValAlaValIlePheAspAlaSerGluIleAspLysVally 408
525 ACCGCGG...CTGTGTATTTAGCCGTTTGACCGCAAGCAAAATCCAG 571
408 sGluGlyAspIleLeuValThrThrThrAsnProAspMetValProA 425
572 TTTGTAGGACGCTGGCGAGAGTGCCTGTGAATATCT..... 612
425 lametLysArgAlaAlaAlaIleValThrAspGluLysGlyArgThrSer 441
613 ...GCCAATGCAACACATGATTTGCGCGCGCGCATCTGCGGTTT 659
442 HisAlaAlaIleValSerArgGluLeuGlyIlePro..... 453
660 GAGTGCAGCAGCATTCATTCATTCAGCGCGGTCGCGGCAATTAACCG 709

```

```

453 ..... 453
710 TGTGACCATCAATTATCAAGATGTAATTACATTGCGCGTTGTTGCA 759
454 ..... CysValValGlyThrLysGluLeu 461
760 ACAGCGCGCTGACACCGAGCGCGTGAATGCCCTAGTGGTTCTCA.. 807
462 ThrLysLysLeuLysThrGlyMetLysValThrValAspGlyThrArgG1 478
808 ...GTCACAACACCGCGCGCTCTTCCGTAACGTTTGGGTGGGAAAGTAT 853
478 yLeuValThrLysGlyLysLeuLysSerLeuValLysLysLysGluGua 495
854 CCCAATTACTGCGGCGAATGTTGACACACACACCGCGGTGTTTC 903
495 lAlaAlaGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 511
904 GGTTCGGTATG..... 915
512 GlyThrMetValLysValAsnValSerMetProGluValAlaGluArgAl 528
916 .....AACGCGCG.....ATTACACACAGCGCGCAGATT 946
528 aAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMetI 545
947 ATTGGAGCGCTACCAATCAGATTTCGCTTATGCAAGAGCGCGCAGC 996
545 lLeuSerLeuGlyGlyHisProLysLysPheLysGlyGlyLysGlu 561
997 AAGAAGCTGTTCGGCTGGTGGCGCGGACCGGCAACAACTCCATC 1046
562 GluGluLeuValGluLysLeuAlaGluGlyLysValAlaAla.. 577
1047 GGTTCACACCTTCGCGCTTCTTCAAAACAACTTTCAGATTTCACAA 1096
578 AlaPheTyrProArgProVal...TrrPyrArgThrLeuAspAlaProTh 593
1097 CAGCGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
593 rAsnGluPheArgGluMetPro..GlyGlyGluAspGluProGluGlu 609
1147 CGCGTGATCCCTTGATATCTCGC.....CACCT 1178
609 rGAsnProMetLeuGlyTrrPArgGlyLysArgGlyLeuAspLInPro 625
1179 GCTTTGGCGGATTAATCGTGGCGGATACCGACGCGCGCGCATTTGG 1228
626 GluLeuLeuArgAlaGluPheLysAlaLeuLysValValGluLysG1 642
1229 GTTGGTGAATTTGGACGAGACACCGCGCTTGTGACAGTGTGCTG. 1277
642 yTrrAsnAsnLysGly.....ValMetLeuProLeu 653
1278 .....CCGCGCAATACGATACGCGCGCGCTTGTGCGCAAGTGTCT 1319
653 aLserHisProGluGluLInLeuArgLysAlaLysArgLInLeuAlaArgGluVal 669
1320 GGA 1322
670 Gly 670

```

seq_name: Swissprot_40:SON_MOUSE

seq_documentation_block:

ID SON_MOUSE STANDARD; PRT; 2404 AA.
AC Q9QX47; Q9QX47; Q9QX47; Q9QX47;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE SON protein.
GN SON.

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Sv;
RX MEDLINE=20408886; PubMed=10950926;
RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
RA Zammit P., Dadrach K., Mazrani W., Kessling A., Lee J.S., Bulwela L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
RT and human genomes.";
RT Genomics 68:57-62(2000).
RN [2]
RP SEQUENCE OF 1-116 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Felschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker G., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
CC -I- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
CC sequence: 5'-GAG[GT]AN[CG][AG]CC-3'. Might protect cells from
CC apoptosis. Might be involved in pre-mRNA splicing (by similarity).
CC -I- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -I- TISSUE SPECIFICITY: Widely expressed.
CC -I- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -I- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, AF193606; AAF23120.1; -
DR EMBL, AF193595; AAF23120.1; JOINED.
DR EMBL, AF193596; AAF23120.1; JOINED.
DR EMBL, AF193597; AAF23120.1; JOINED.
DR EMBL, AF193598; AAF23120.1; JOINED.
DR EMBL, AF193599; AAF23120.1; JOINED.
DR EMBL, AF193600; AAF23120.1; JOINED.
DR EMBL, AF193601; AAF23120.1; JOINED.
DR EMBL, AF193602; AAF23120.1; JOINED.
DR EMBL, AF193603; AAF23120.1; JOINED.
DR EMBL, AF193604; AAF23120.1; JOINED.
DR EMBL, AF193605; AAF23120.1; JOINED.
DR EMBL, AF193607; AAF23120.1; JOINED.
DR EMBL, AK019312; BAB31656.1; -
DR EMBL, AK019081; BAB31536.1; -
DR EMBL, AK008478; BAB25691.1; -

```

DR	EMBL: AK008256; BAB25562.1; -.
DR	MGD; MGI:98353; SON.
DR	InterPro: IPR001159; DS_RBD.
DR	InterPro: IPR000467; G_patch.
DR	Pfam: PF00035; dsrm; 1.
DR	Pfam: PF01585; G_patch; 1.
DR	SMART: SM00443; G_patch; 1.
DR	PROSITE: PSS0137; DS_RBD; 1.
DR	PROSITE: PSS0174; G_PATCH; 1.
KW	RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW	Alternative splicing.
FT	DOMAIN 721 850
FT	687 943
FT	DOMAIN 961 1080
FT	REPEAT 961 966
FT	REPEAT 969 974
FT	REPEAT 976 981
FT	REPEAT 985 990
FT	REPEAT 993 998
FT	REPEAT 1001 1006
FT	REPEAT 1010 1015
FT	REPEAT 1018 1023
FT	REPEAT 1026 1031
FT	REPEAT 1035 1040
FT	REPEAT 1044 1049
FT	REPEAT 1055 1060
FT	REPEAT 1066 1071
FT	REPEAT 1075 1080
FT	DOMAIN 1101 1133
FT	1910 1979
FT	REPEAT 1910 1916
FT	REPEAT 1938 1944
FT	REPEAT 1945 1951
FT	REPEAT 1952 1958
FT	REPEAT 1959 1965
FT	REPEAT 1966 1972
FT	REPEAT 1973 1979
FT	DOMAIN 1919 1990
FT	REPEAT 1919 1937
FT	REPEAT 1980 1990
FT	DOMAIN 1991 2017
FT	2283 2329
FT	DOMAIN 2349 2404
FT	VASAPLIC 2086 2086
FT	VASAPLIC 2087 2404
SO	2404 AA; 261428 MW; 448B528BD3FC01D9 CR664;

```

Alignment_scores:
      Quality: 105.50      Length: 506
      Ratio: 0.434      Gaps: 24
Percent Similarity: 48.024      Percent Identity: 23.123

Alignment_block:
US-09-303-518D-125 x SON_MOUSE ..

Align seg 1/1 to: SON_MOUSE from: 1 to: 2404

18 AGGCTTAAACCTCGCCATCGCGGGAGACCGGACGACG.....CG 58
|||||..... |||||.....: |||||..... ||
1860 ArgSerArgSerIysSerIysSerIysArgIysArgIysSerValSerLysGluLysArg 1876
59 TTTCAGAGGGCCCGCCCATACCGAAGTGCCTGCTGGCGAGAAATAT 108
||||| ||| ||| ||| |||||.....: |||||.....
1876 GluLysArgSerProLysHis...ArgSerIysSerArgIysArgLysArgL 1892
109 GCCGGTATGCGCCCTTCGATGAAATCAAGCAAGCAAGCCGATCCGCTCAAAA 158
||| ... |||||.....: |||||.....: |||||.....: |||||.....

```

```

894 CGGATTTCCGGTGGATTTAGACGGCGGCGGATTACAGAGCGCGCAG 943
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2188 nvalpheserleuprosergluproval...asp.ilserthrl 2203
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
944 ATTATTTGGACGCTACCAATCAGATTTCCGTTATGAGAAAGCGCGC 993
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2204 Almetsergluarvalaleuvalaglnysargleusegluasnalaph 2220
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 AGCAAGAGGCTGTGGCTGGTGGCGCGCGGACCGGACGCAATACCTCAT 1043
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2220 eaapleuglualametermetleuasnargalaglnlunarglileaspa 2237
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 CACGCGTACACCGTGGCGGATTTCTGTAACAACT...CTTCA 1087
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2237 latrpalaglnleuasnserile.ProgylglnPheThrGlySerThrgl 2253
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 AGTTCAACACAGCGGTCACGCGGCGGCGGCGCATGTGTGC... 1130
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2253 yvalglvalleuthrglnleuvalasnthrglvalaglnalat 2270
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1131 .....GATTGGTACTTACGAGCGCGTGATGCCCTTGATATCTCT 1169
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2270 rpllethylsaspalnphelenuargalalaprovalthrgly... 2284
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1170 GCCCACCCTCTTTGGCGGATTTAATCGCGCGGATACCGACGCGCGC 1219
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2285 ...glymetglyalvalleumetargylsmetglytrpargluylgl 2300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1220 AGGATTTGGGTTGCTTGAATTGACGAGAGACCTCCCTTGTGTCAGC 1269
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2300 ucllyuclglyllysnlnlysglnlyasnlysglnproulileuvalasp 2317
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1270 TTCCTCTG.....CCCGGCAATATACGATACGCGCGCTGTGGCGCA 1313
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2317 helysthrasparglyglyleuvalalvalglygluargalaglnlys 2333
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1314 AGTGCTGGAAC 1325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2334 Argserglyasn 2337
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: SwissProt_40:MURF_HAEIN
seq_documentation_block:
ID MURF_HAEIN STANDARD; PRT; 457 AA.
AC P45061;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoylalanine-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNac-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
CN MURF OR H11134.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischiemann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uettermack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RA influenzae Rd."
RL Science 269:496-512(1995).

```

```

CC -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYZES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U32793; AAC22789.1; -.
CC DR HSSP: P11880; 1G94.
CC DR TIGR: H11134; -.
CC DR InterPro: IPR000713; Mur_Ligase.
CC DR InterPro: IPR004101; Mur_Ligase.C.
CC DR Pfam: PF01225; Mur_Ligase.1.
CC DR Pfam: PF02875; Mur_Ligase.C.1.
CC DR Peptidoglycan synthesis: Cell division; Cell wall; Ligase;
CC ATP-binding; Complete proteome.
CC NP_BIND 109 115 ATP (POTENTIAL).
CC FT SEQUENCE 457 AA; 50073 MW; 095FEF74F27CD208 CRC64;

```

```

alignment_scores:
Quality: 105.00 Length: 381
Ratio: 0.550 Gaps: 20
Percent Similarity: 50.131 Percent Identity: 21.785

```

alignment_block:

US-09-303-518D-125 x MURF_HAEIN ..

```

Align seg 1/1 to: MURF_HAEIN from: 1 to: 457
253 CGGCTACTTGCAGTCAGTCGATTTGCGGATTGAGACGACGCAATGCA 302
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      11 GlnleuglnAlalysleulle.....GlyAspGlnasvalgl 24
303 GTTGAACGCTACGCGACCGTGAAGCGCTGGCAACTTAAGCGGGAAGAG 352
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      24 nvalglulysleasnthrpsphr..... 32
353 TGGCGCGACCTGATTCGATCCGGTTTGGGACCTGCGCGGCGACCCGT 402
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      33 .....ArglySerValSerAsnSerleuphehaleuylsrglylu 47
403 CCGTTCACCAAAATTCCTGCCGTCGAT.....GCCAGCGCTTGGCCAT 446
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      48 LysheaspralalsglnlryleuaspslAlalValserglnglAlale 64
447 CTGCTCAATGCGATGAGACACCAATCCGCTGCGTCCGACCCCTACGCTCA 496
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      64 uAlaleuvalValglnglnlgnlgnlgnlgnlgnlgnlgnlgnlgnl 81
497 TTATCAAGAA.....CGCGCGAGATTTCAACGCGCGCTGTG 537
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      81 alVallyaspThrArglleAlaleuylgluueuAlalyslrypleu... 96
538 GTATTGACCGCTTGACCGCAAGCAAAATCCATGTTTGTAGCAGCTGG 587
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      97 ...ArggluylsleasnproargThrValAlamethrglyserSerl 112
588 CGCACACGTGCGGCTGTAATGCTGCGACATGGAACATGAAATTCG 637
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

0G Chloroplast.
0C Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
0X Actinoptoraceae; Pyraliella.
0N NCBI_TaxId=2895;
0P [1]
0R SEQUENCE FROM N.A.
0X MEDLINE=92216062; PubMed=153750;
0A Jouanne S., Ketrourc H C., Kloarec B., Loiseaux-De Goer S.:
0T Nucleotide sequences of the atpB and atpE genes of the brown
0T alga Pyraliella littoralis (L.) Kjellm. ";
0L Plant Mol. Biol. 18:819-822(1992).
0C -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
0C GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
0C SUBUNIT.
0C -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
0C H(+) (Out).
0C -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
0C CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
0C SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
0C HAS THREE MAIN SUBUNITS: A, B AND C.
0C -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
0C -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
0C
0C -----
0C This SWISS-PROT entry is copyright. It is produced through a collaboration
0C between the Swiss Institute of Bioinformatics and the EMBL outstation -
0C the European Bioinformatics Institute. There are no restrictions on its
0C use by non-profit institutions as long as its content is in no way
0C modified and this statement is not removed. Usage by and for commercial
0C entities requires a license agreement (see http://www.isb-sib.ch/announce/
0C or send an email to license@isb-sib.ch).
0C

```

```

204 TACTGGCGGGCTTCAGGCAATCCCGGATTCACCTGGCGGAAAGC 253
68 tSerAla.....ThraspIyleuGlnaIrgIyValThyV 80
254 GCGTACTGACGTCAGTGGATTGCGGTGA..... 285
80 aIIleasPrhIrnAlaProIIeAlaValaIroValaIySaIaThIleu 96
285 ..... 285
97 GIIATrGlIePhaSnValLeuGlyIlnProValaIAspRnleuSerIpsE 113
286 ....GGCAAGCAGCAAAATCGATTGGACGCTAGCAGCCTGAAGCGCTG 331
113 rValGlyIlnAspPrhIleuProIIeIlnAsrSerAlaPro...AlaPhet 129
332 CAACCTTAAGCGGGAAGAGTGGCGGCAAGCTGATCCAAACCGGTTG 381
129 hrAspLeu.....GluThIysProAlaIlePhaGluThIyIle 142
382 TGGACCTGCGTGGCAACCCGCTTCAGCAAAATTCCTGCGTGGATGC 431
143 LysValValaIAspRnleuAlaProIyArGArg.....G1 154
432 CGAGCCGTCGCAATCTTCGTCATCGCATGACACCAATCCGCTGGTG 481
154 yGlyIyIleGlyLeuPhaGlyIyAla.....GlyV 165
482 CCGACCTAGGTCATTCATCAAGAGCCGCGAGATTCATCAACGC... 528
165 aIlglyIsthrValleuIIeMerGluIleuIleasnIIeAlaIySala 181
529 .....GCGCTGTGTATTGAGCCGTTTGAACGACGCAAAATCCATGT 572
182 HisGlyIyValSerValaPhaGlyIyAlaIyGlyIyArG.....194
573 TTGTAAAGCACTGGCGGACGCGCGCTCGAATAATGTGCCAACATCG 622
195 .....ThraIrgGlyIyAsnAspLeuIyMetG 204
623 AAACATGATTCGGCGGCGCGCATCCGCGGTTTGAAGTGCACGAC 672
204 IuMetIySergIyValIleasnGluThraIlnleuGluIySerIyS 220
673 ATTCATTTCATC.....GAGCGGTCGCGCGGATTAAC 707
221 ValAlaIeValaIyrgIyGluMetAsnGluProGluIlaIaIyMetAr 237
708 CGTGTG.....ACATCAAT 724
237 gValGlyIeuthraIleuthraIleuIaIgluTyIyPhaIrgAspIleasnI 254
725 ATCAAGATGATTAATAC...ATTGGCGCTGTGTTCACAGCGCGCTG 771
254 ySgIIaIyValleuIeuthraIleuIaIlePhaIrgPhaValaIeIlna 270
772 AACACGAGCGCGTATTGCCCTAGT.....GGTCTCAAGTCAACAA 815
271 GlySerGluValSerAlaIleuIeulIyArGmetProSerAlaValaIgluTy 287
816 ACCGCGCTTGGTACCGTTTGGTGGCAAGATACCAATTAATG 865
287 rGlnProThIleuGlyThIrgIuMetGlyAlaIeulnGluIyIleThr 303
866 CGGGGCAATGGTTGACACAGCAACCGCGTATTCGGTTCCGTTATG 915
304 .....SerthrThr 306
916 AACGGCGCATTAACAAGCGCGGACAGATTAATTTGGACGCTACCA 965
307 GlnGlySerIleThrSerIleGlnaIaValaIyIyVal..... 318

```

```

966 TCAGATTTCGCTTATCGAAGAGCGCGCAAAAGACCTGTGGCGG 1015
319 .....ProIIaIAspRnleu.....T 324
1016 TTGGCGCGGACCGGACAAATCTCATCAAGCTGACACCTGGCGCAT 1065
324 hrAspRnIaIPro.....AlaThIrhIrhPhaIlaIlnI 324
1066 TTCCSTGAATAAACAAGCTTCAGATTCAACAGACGCTCAACGGCGG 1115
335 Leu.....AspAlaThIrhIyValleu 342
1116 CCGGCGCATGGTGGCGGATTCGCTACTACAGCGCGGTATGCGCTTGATA 1165
342 rArgGlyIleuAlaIaIySgIyIleIyrrProIIaIValaIAspRnleu 359
1166 TCTGCGCCGACCTGCTTTGGCGGATTAATGTCGCGCAT..... 1206
359 erThrSerThr...MetIeulnProIleuIleValaIyAspRnIlnIstYr 374
1207 .....AC 1208
375 LysThrAlaGlnleuValaIySgIuThIleuIlnArGlyIySgIuIeul 391
1209 CGACAGCGCGGACGATTCGCTTCGTAATGGACAGCAAGAC... 1254
391 nAspIleIleAlaIleuGlyIleAspGluIleuSerGluIlnAspRgl 408
1255 .....CTGCTTTGTCAGC 1269
408 euValaIAspRglAlaArGlyIleGluArGhIeIeulSerGluProPha 424
1270 TTGCTGTGC.....CCGGGCAAAATC 1290
425 rHeValaIaIgluValaIrhThrcIySerProGlyIyStyr 437
seq_name: SwissProt_40:PPSA_PYROHO
seq_documentation_block:
ID PPSA_PYROHO STANDARD; PRT; 821 AA.
AC 057830;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable phosphoenolpyruvate synthase (Ec 2.7.9.2) (Pyruvate, water
DE dikinase) (PEP synthase).
GN PPSA OR PH0092.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=96344137; PubMed=9679194;
RA Kawarayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Toshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -! CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC phosphoenolpyruvate + phosphate.
CC -! PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC LACTATE ARE USED AS A CARBON SOURCE.
CC -! SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).

CC -----
 DR EMBL: AP000001: BAA29161.1; -
 DR HSSP: p22983; IDIK.
 DR InterPro: IPR000121; PEP_utilizers.
 DR InterPro: IPR002192; PPK_N_term.
 DR Pfam: PF00391; PEP-utilizers; 1.
 DR Pfam: PF02896; PEP-utilizers; 1.
 DR Pfam: PF01326; PPK_N_term; 1.
 DR ProDom: PD000940; PEP_utilizers; 1.
 DR PROSITE: PS00370; PEP-utilizers; 1.
 DR PROSITE: PS00742; PEP-ENZYMES_2; 1.
 DR Transferrase: Kinase; ATP-binding; Phosphorylation; Complete proteome.
 KW MOD_RES 444 444 PHOSPHORYLATION (BY SIMILARITY).
 FT DOMAIN 812 818 POLY-GLU
 SO SEQUENCE 821 AA; 90812 MW; A93816D865F8A0BF CRC64;

alignment_scores:
 Quality: 104.50 Length: 455
 Ratio: 0.437 Gaps: 21
 Percent Similarity: 52.527 Percent Identity: 21.099

alignment block:

US-09-303-518D-125 x PPSA_PYRHO ..

Align seq 1/1 to: PPSA_PYRHO from: 1 to: 821

```

67 GCGCCGCGCATTTACGAGAGTCGCTGCTGGCGAAGATATGCGGAT 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 GYGLIALAVALAVALSERGLAVALTPROASPGUTRYLLEVAL 271
117 GCGCCGCGCATTTACGAGAGTCGCTGCTGGCGAAGATATGCGGAT 166
   :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 uLysgLTThrTPrLysLTLeLysGLu...LysValLTLeAlaLysLTLeLysGLu 287
167 TGCTGTTGAGACAAAAGATCCGGCGGCTGTTACTGCGCGGCT 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 alMet.....ValLTLeArgAsnProLU 294
217 TCAGGCAAA.....ATCGCGGATTCACCGTGC..... 246
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 ThngLTyLsgLTThrValGLInValLysValAlaGLUTRYLeuGLYProGL 311
247 .....GAAAGCGCGTACTT.....CACTCACTGCTGATTCGCGT 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 uTPValGLuLysGLInValLTThrAspGLuGLInLTLeGLuValAlaL 328
284 AAGGACAGACGAATGAGTTTGACGCTACGCA...CTGAAGCGCTG 330
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 yMeLsgLTyLsgLTLeGLu...GLuHsTYRGLYTPProGLInAspLe 343
331 GCA...AACTTAAGCGCGCAAGAGTGGCGCAACTGATTCGCAATCGG 377
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 GLUTPAlaTYRAspLYAspAspGLYLYsLeuTYRLeValGLInSerAT 360
378 TTTGAGACTGCGTGGCGCACCGCTCCGTTACGAAATTCCTGCGCTG 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 gProLTLeThrThrLeuLysGLuThrThrThrGLuGLuValGL 376
428 ATGCGGCGCGTTCGCTGCTGCAATGAGTACGACCACTGCG... 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 .....GLuGLuAlaGLuValLTLeuLysGLYLeuGLYAlaSerProGLY 391
475 CTGGCTGCGGACCTACGCTCATTAATCAAGAAGCGCGGAGGATTTCAA 524
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 LTLeGLYAlaGLYArgValAlaValLTLePheAspAlaSerGLu..... 405
525 ACGGCGCTGTTGATTAAGCGGTTTGACGCAAGCAACCAATTCATGTT 574
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406 .....LTLeAspLYsValLYsGLuGLYAspValLeuValT 417

```

```

575 GTAACGACGCTGGCGAC...GTGCGCT...GAAATGCTGCCAAC 618
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 hrThMetThrAsnProAspMetValProAlaMetLysArgAlaSerAla 433
619 ATCGAAACATGAATTCGGGCGCGCCGATTCGCTGCTGGTGGAC 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
434 LTLeLTThrAspGLuGLYArgThrSerHsAlaAlaLTLeValSerAT 450
669 GCACATTCATTCATTCGACCGCGCTGGCGGCAATTAACCGTGGACCA 718
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 gLInLeuGLYLTLeProAlaValAlaGLYThrLysGLuAlaThrLysGL 467
719 TCAATTAACAAGATGAATTAACATTCGCTGCTGCTGCTGCTGCT 747
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 euLysThrGLYAspTYRValThrValAspGLYThrArgGLYLeuValTYR 483
748 .....CGTTGTTGCAACAGCGCGCTGACACCGGCGCGCT 785
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 LysGLYLTLeValLysSerLeuValGLYLYsLYsGLuGLuAlaL 500
786 GATTGCGCTGAGTGTTCGATTCGACGCAACGCGCGCTGCTGCTGCTG 835
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 aAlaAlaProGLYAlaAlaValAlaAlaAlaProLeuValThrGLYThrL 517
836 TTTTGGCTGCGAAAGTATCGCAATTAACGCGGCGGCAATGCTGTCAC 885
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 euValLYsValAsnValSerMetProGLuValAlaGLu..... 529
886 GACACCGCGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 .....ArgAlaAlaAlaThrGLYAlaAspGLYValGLYLeuValArg 544
936 GCGGACGATTAATTTGGAGCTTACACACATGATTCGCTGCTGCTGCTG 985
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 aGLuHsMetLTLeLeuSerLTLeGLYGLnHsProValLysPheLTLeLysG 561
986 AAGCGCGGCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1035
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
561 LTLeLYsGLuAspGLuLeuValGLuArgLTLeGLuGLYLTLeGLuLys 577
1036 TACTCATCAGCGCTACACCGCTGCGGATTCCTGAAACCAACTCTT 1085
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
578 ValAlaAla..AlaPheTYRProArgProVal...TPRYArgThrLeu 592
1086 CAAGTTCAACACGCGCTCAACGCGGCGGCGGCGGCGGCTGCTGCTG 1135
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
593 AspAlaProThrAsnGLuPheArgGLuMetPro...GLYGLYGLuAspGL 608
1136 GTACTTACGAGCGCGCTGATGCCCTTGATATCTGCC..... 1172
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
608 uProGLuLTLeuArgAsnProMetLTLeuGLYTPArgGLYLTLeArgArgLYL 625
1173 .....CAOCTGCTTTTGGCGGATTTATGCTGCGGCAATTCGCGGCGG 1217
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
625 euAspGLnProGLuLeuLeuArgAlaGLuThrLysAlaLTLeLysVal 641
1218 GCAGGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
642 ValGLuLysGLYTYRAsnAsnLTLeGLY.....ValMe 652
1268 GCTTGCCTGCTG.....CCGCGGCAATTAAGATATACGCGCGCTGCTG 1308
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
652 tLeuProLeuValSerHsProGLuGLnLTLeArgLYsAlaLYsGLuLTLe 669
1309 CGCAAAATGCTGGA 1322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
669 LTLeArgGLuValGLY 673

```

seq_name: SwissProt_40:FIG2_YEAST

seq_documentation_block:

ID FIG2_YEAST STANDARD: PRT: 1609 AA.


```

1305 .ThrsSerSerThnLeuProThraLeuAlaValSerSerThrPhet 1321
127 TCGAGGGGGCATACCGCATATTCT.....TCGCGA 96
1321 LeuLaserSerLeuProLeuSerSerLysSerSerLeuSerPro 1337
95 ACCAACGGCAGCTTCGTAATG.....GCCGGCGCGTCTGA 61
1338 ValSerSerSerLeuLeuMetSerClnPheserSerSerSerSe 1354
60 AACGGCTTCCTCCGCTTCGCCCGCATG 33
1354 rSerSerLeuLaserLeuProSerLeu 1363

```

seq_name: SwissProt_40:ATPB_DICDH

seq_documentation_block:

ID ATPB_DICDH STANDARD; PRT; 481 AA.

AC P30158;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE ATP synthase beta chain (EC 3.6.3.14).

GN ATPB.

OS Dictyota dichotoma.

OG Chloroplast.

OC Eukaryota; stramenopiles; Phaeophyceae; Dictyotales; Dictyotaceae;

OC Dictyota.

NCBI_TaxID=2876;

RP SEQUENCE FROM N.A.

RA MEDLINE-92322956; PubMed-1535802;

RT Leitsch C.E.W., Kowalik K.V.;

RT "Nucleotide sequence and phylogenetic implication of the ATPase

subunits beta and epsilon encoded in the chloroplast genome of the

Plant Mol. Biol. 19:289-296 (1992).

RL FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON

GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC

SUBUNIT.

CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) - ADP + phosphate +

H(+)(out).

CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC

CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE

SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)

HAS THREE MAIN SUBUNITS: A, B AND C.

CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.

CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL: X66939; CAA47370.1; -

DR PIR: S22509; S22509.

DR HSSP: P00829; IBMF.

DR Mendel: 2225; DICDH:atpb.1.

DR InterPro: IPR004100; ATP-synt_ab_N.

DR InterPro: IPR000793; ATPase_AB_C.

DR InterPro: IPR000194; ATPase_alpha_beta.

DR Pfam: PF00006; ATP-synt_ab_1.

DR Pfam: PF00306; ATP-synt_ab_1.

DR Pfam: PF02874; ATP-synt_ab_C_1.

DR PROSITE: PS00152; ATPASE_ALPHA_BETA_1.

DR ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);

Hydrolysis; ATP-binding; Hydrogen ion transport.

NP_BIND 161 168 ATP (BY SIMILARITY).

SEQUENCE 481 AA: 52203 MW: 849515C113FD1F91 CRC64;

alignment_scores:

Quality: 103.50 Length: 515
Ratio: 0.502 Gaps: 25
Percent Similarity: 40.000 Percent Identity: 21.942

alignment_block:

us-09-303-518d-125 x ATPB_DICDH ..

Align seg 1/1 to: ATPB_DICDH from: 1 to: 481

```

67 GCGCCGGCCATTAACGAGTCGCTTCCTTCGCGAAGATATGCGGTAT 116
|||||
21 GlyProValIleAspAlaValPhe.....SerIleGlyI 32
117 GCGGCC.....TCGATGAAGTCAGAGGCGATGCC.... 150
: ||| :
32 nleuProLysIleTyraSnaIleuGluValLysSerLysAspGlyThrT 49
151 .....GTCAAAAAGCGCAAGTGTCTTGAACACAAAAGATCCGGCC 195
: : : ||| :
49 hrlleIleCysGluValGlnGlnleuPheSnaSpsnIleValArgAla 65
196 GTGCTGTTACTCGCGCGCGCTTCAGCAAAATCGCGCATTCACCGTGG 245
: : : : :
66 IleIleMetSerIle.....ThrAspGlyLeuGlnArgI 77
246 CGAAG..... 252
77 YValGluValIleAspThrGlnAlaProIleLeuValProValGlyLysA 94
253 .....CGGTACTTCAGCACTGCTGATTCGGCTT 282
: : : : :
94 latrIleuGluArgIlePheAsnValLeuGlyGlnThrValAspSnaIle 110
283 GAA.....GCAACAGCAAGATTCAGTTGAACGCTACGACCTGA 323
||| :
111 GlnIleGlyThrGlyLysAspArgLeuProIleHisArgProAlaProSe 127
324 AGCGCTGCGCAACTGAAGCGCGAAGAGTGGCGCGCAACCTGATCAAT 373
:
127 r..... 127
374 CCGGTTTGAGACTCGCTGCCACCGCTTCGCGAAGATTCCTGCC 423
: : : : :
128 .....PheThrAspLeuGluThrLysPro..... 135
424 GTTCGATGCGAGCGCTTCGCGCATTCGCAATGCGATGCGACCAATCC 473
136 ..... 146
474 GCTGCTGCC.....AlaIlePheGluThrGlyIleValValAs 146
||| :
146 PheIleuAlaProTyArgArgGlyGlyLysIleGlyLeuPheGlyLysA 163
484 .....GACCTACGTCATTAATCAAGAAAGCGCGAGAGATTGAA 525
||| :
163 IagIleValGlyLysThrValIleuIleMetGluIleAsnIleAla 179
526 CCG.....GCGCTGTGTATTAAGCGCTTTCGACGCGAAGCAAAAT 566
: : : : :
180 LysAlaHisGlyLysValIleSerValIlePheGlyGlyValGlyValArg 194
567 CCAATGTTGTGAAGCGAGCTGCGCGAGACGTCGCTGAAAATGCTGCCA 616
: : : : :
195 .....ThrArgGluGlyAsnAspLeu 202
617 ACATGGAACACATGAATTCGCGCGCGCGCGATTCGCGGTGAGATGG 666
: : : : :
202 yrmGluMetLysGluSerGlyValIleAsnGluSerAsnLeuSerGlu 218
667 ACGCAATTCATTTCATC.....GAGCGGCTGCGCGCGAA 701

```

```

::: ::: ::::: ||||| |||||:
219 SerLysValAlaLeuValTyrGlyGlnMetAsnGlnProGlyAlaLar 235
702 TAAACCGTGG.....ACCA 718
235 gmetatgValGlyLeuThrAlaLeuThrMetAlaGluTyrPheAlaGAspI 252
719 TCATATTCAGATGTAATTACC...ATTGGCCGTTTGGTTCACACAGCC 765
252 IeAsnArgGlnAspValLeuLeuPheIleAspAsnIlePheArgPheVal 268
766 CGTGTGACACCGGCGCGTGGTGGCCCTAGT.....GGTTCACAGT 809
269 GlnAlaGlySerGluValSerAlaLeuLeuGlyArgMetProSerAlaVala 285
810 CACAAACCGGCGCTTGGCTACCGGTTTGGTGGCAAGATACGCAAA 859
285 IGlTyrGlnProThrLeuGlyThrGlnMetGlyAlaLeuGlnGluArgI 302
860 TTACTGGGGCGAATTGGTTGACACAGACACCGCGTTCGGTTCG 909
302 IeThr.....Ser 304
910 GATATGACGCGCGGATTACACAGCGCGCAGATTATTTGGACGCTA 959
305 ThrTrgInGlySerIleThrSerIleGlnAlaValTyrVal..... 318
960 CCACAAATCAGATTCGTTATCGAAGAGCGCGACAAAGACGCTGCG 1009
319 .....ProAlaAspAspLeu.... 323
1010 GCTGGGTGGCGCGCGGCAAAATATCTCATCAGCGCTACACCCCTC 1059
324 .....ThiAspProAlaPro.....AlaThrThrPhe 332
1060 GGCCATTTCTGAAACAACTCTCAAGTTCAACACAGCCGTCACGG 1109
333 AlaHisIleu.....AspAlaThrThrVa 340
1110 CGGCGACGCGCGCATGTCGCGATTGACTTACGACGCGGTATGCCCT 1159
340 ILeuSerArgGlyLeuAlaAlaLysGlyIleTyrProAlaAlaAspProL 357
1160 TGGATATCTGCGCCACCTGCTTGGCGGATTTATCGTGGCGGATACC 1209
357 euAspSerThrSerThr...MetLeuGlnProValIleValIleSerGlu 372
1210 .....GACAGCGCGCAG..... 1221
373 HisTyrAspThrAlaGlnLeuValLysLysThrLeuGlnArgTyrLysG1 389
1222 .....GCATTGGGTTGGCTGGATTTGGACGACAGAG 1252
389 uLeuGlnAspIleIleAlaIleLeuGlyIleAspGlnLeuSerGlnGlu 406
1253 AC.....CTGCGCTTGG 1263
406 spArgLeuValAlaAspArgAlaArgLysIleGluArgPheLeuSerGln 422
1264 TGCAGCTTGCTGCG.....CGGGCAAAATAC 1290
423 ProPhePheValAlaGlnValPheThrGlySerProGlyLysTyr 437
seq_name: SwissProt_40:VL2_HPV31
seq_documentation_block:
ID VL2_HPV31 STANDARD: PRT: 466 AA.
AC P17389;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Minor capsid protein L2.
GN L2.

```

```

OS Human papillomavirus type 31.
OC Viruses; dsDNA viruses; no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299478; PubMed=2545036;
RA Goldsborough M.D.; Dislivestre D.; Temple G.F.; Lorincz A.T.;
RT "Nucleotide sequence of human papillomavirus type 31: a cervical
RT neoplasia-associated virus."
RL Virology 171:306-311(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: J04353; AAA6955.1; -.
CC PIR: H32444; P2WL31.
CC InterPro: IPR000784; late_L2.
CC Pfam: PF00513; late_protein_L2; 1.
CC Coar protein; Late protein.
KW SEQUENCE 466 AA; 49975 MW; 99AD125BC65B1368 CRC64;
SQ

```

```

alignment_scores:
Quality: 102.50 Length: 368
Ratio: 0.614 Gaps: 21
Percent Similarity: 45.380 Percent Identity: 23.641

```

```

alignment_block:
US-09-303-518D-125 x VL2_HPV31 ..

```

```

Align seg 1/1 to: VL2_HPV31 from: 1 to: 466

```

```

354 GCGCGCACCTGATCCAAATCCGTTGTGACTGGCGCTCGCACCGCTC 403
119 AlaProAlaProIlePro.....HisPro_P 127
404 CG.....TTCAGCAAAATCTCTGCGCGTATGCGGACCGCTTC 441
127 rothThrSerGlyPheAspIleAlaThrThrAlaAspThrThrProAla 143
442 GCCATCTTGTCAATGCGAGGACAC.....AATCGCTGGCTGGCA 485
144 IleuAspValThrSerValSerThrHisGlnAsnProThrPheThrAs 160
486 CCTACGCTCATATCAAGAAAGACCGCGGAGATTTCACAGCGGCGCT 535
160 ProSerValLeuGlnProProThrProAlaGlnThrSerGlyHisLeuL 177
536 TGGTATTAAGCGGTTTGACCGACGCAAAATCATGTTTGTAAAGCAGCT 585
177 euLeuSerSer.....SerSerIleSerThrHisAsnTyrGlnGluIle 191
586 GCGCGCAGAC.....GTCCGCTGAAATGCTCCCAACATC.....GA 623
192 PrometAspThrPheIleValSerThrAsnAsnGlnIleThrSerSe 208
624 AACACATGAATTCGGCGCGCGCATCTCGCGGTTTG..... 660
208 ThrProIleProGlyValArgArgProAlaArgLeuGlyLeuTyrSerL 225
661 ..AGTGGCAGCATTCATTCATCGACCG..... 690
225 ysaIaThrGlnGlnValLysValIleAspProThrPheLeuSerAlaPro 241
691 .....GTGCGCGGAA 701
242 LysGlnLeuIleThrTyrGlnAsnProAlaTyrGlnThrValAsnAlaG1 258

```

```

702 TAAACCGGTGTGACCATCAT..... 723
258 ucGuseurleuTyrrheserAsnThSerHisAsnIleAlaProAspProA 275
724 ..TATCAAGATGTAATACATATGCGCTGTGTTCAGACGGCGGCTG 771
275 sPpHeuAspIleIleAlaLeuHisAsnProAlaLeuThSerArg 291
772 AACACCGGCGGTGATGCGCTAGTGGTTCAGTCAACAACCGCG 821
292 AsnThrValArgTySerArgLeuGlyAsnGlySlnThrLeuArgThrAr 308
822 CCTCTTGGGACCGCTTGTGGCGCAAGTA.....T 853
308 gserGlyAlaThrIle..GlyAlaArgValHisTyTyTyAspIles 324
854 CGCAATATCT..GCGGCGAATGTGTGACACAGCAACCGCGTAT 900
324 erSerIleAsnProAlaGlyIleSerIleGluMetGlnProLeuGlyAla 340
901 TCCGTTGCGTATGACGGCGCATACAGCGCGCGACATATTT 950
341 SerIaThrThr..ThrSerThrLeuAsnAspGlyLeuTyAsp..IleT 356
951 GCGACGCTAC.....ACAATCAGATTTCCG.....TATCG 982
356 yAlaAspThrAspPheThrValAspThrProAlaThrHisAsnValSer 372
983 AAGAAGCGCGCAGCAAGACCTGTGCGGGGCGTCCGCGCGCGCGAC 1032
373 ProSerThrAlaValGlnSerThrSerAlaValSerAlaTyValProTh 389
1033 AAATACT..... 1039
389 rAsnThrThrValProLeuSerThrGlyPheAspIleProIlePheSerG 406
1040 .....CCATCAGCGCTACACCGCTCGGCC 1063
406 LyProAspValProIleGluHisAlaProThrGlnValPheProPhePro 422
1064 ATTTCTGAAAAACAACCTTCAGTTCACACAGCGCGTCAAGCGCGC 1113
423 Ieu.....AlaProThrThrProGlnValSerI 432
1114 GACCGGCGCATGTCGCGCATGTGACTTACGAGCGGTGATCGCTTGA 1163
432 ePheVal.AsplGlyAspPheTyIeu..... 441
1164 TATCTGCCCGACCTGCTTTCGCGATTTATGTCGGCGCATACCGA 1211
442 .....HisProSerTyTyTyMetLeuLysArgArgTyLysArg 454
seq_name: SwissProt_40:RNFC_PASMU
seq_documentation_block:
ID RNFC_PASMU STANDARD; PRT; 835 AA.
AC Q9CNP2;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfC.
GN RNFC OR PM0385.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxId=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT *Complete genomic sequence of Pasteurella multocida pm70.*;

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4Fe-4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; AE006074; AK02469.1; -
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR001949; Complex1_51K.
DR Pfam; PF01512; Complex1_51K; 1.
DR ProSITE; PS00037; fer4; 2.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 835 AA; 90851 MW; E682CAAF7A19E79 CRC64;

alignment_scores:
Quality: 102.50 Length: 447
Ratio: 0.477 Gaps: 18
Percent Similarity: 48.098 Percent Identity: 19.463

alignment_block:
us-09-303-518d-125 x RNFC_PASMU ..
Align seg 1/1 to: RNFC_PASMU from: 1 to: 835

103 GAATATGCGCGTATGCGCCCTCATGAAGCAAGGCGATGCCGT 152
::::::::::: |||||
48 GlnHisAlaGlyGluAlaGlyAsnIleLeuValLysValGlyAspTyVal 64
153 CAAAAAGGCCCAAGTCGCTGTTGAAGACAAAAAGAACCGCGGTGTGT 202
||||| |||
64 IPhelysGlyGlnProLeuThrGlnGlyAspGlyLeuArgValLeuProV 81
203 TTAATGCGCGCGCTTCAGCAAAATCGCGCGAT.....CACCGT 243
||| ::||| ||| |||||
81 AlnHisAlaSerThrSerGlyPheIleArgAlaIleAlaProTyAlaSer 97
244 GCGGAAAAAGCGCGTCTTCAGTCAGTCGATTCGCTTGAAGCAACGA 293
::::: ||| :::::
98 AlnHisProSerGlyLeuAlaThrLeuLysLeuHisIleGluLysAla 114
294 GCAATCGAGTTTGACGCTACGACCT.....GAAGCGCTGCAAACTTAA 340
::: ::::: ||| :::::
114 yLysAspGlnThrArgGlnGlnGlnProLeuAspAspIleLeuThrGln 131
341 GCGGGAAGAAAGTCCGCGCAACCTGATCCATTCGGTTGTGACTGCG 390
::: ||::: ||| ||::: |||
131 hrProGlyLysLeuIleGlyLysLeuTyGlnIleGlyVal..AlaGly 146
391 CTGCGCAACCGCTCGCTTC.....AGCAAAATTCCTGCGCTGATGC 431
||| ::||| :::::

```

```

147 LeuGIyGIyAlaValPheProThrAlaAlaLysLeuHisSerAlaGluLy 163
432 CGAGCGCTGGCCATCTTCATCGATGAGACACATCCGCTGGCTG 481
163 scinValLysLeuLeuIleIleasnGlyAlaGluLysGluProTyrIleT 180
482 CGGACCCCTAGCTCATTCATCAAGAAGCCCGGAGATTCAAAAGCGGC 531
180 hrcysAspAspArgLeuMetArgSerPtyrAlaAspGluIleLeuGlu 196
532 CTGTGGTATGAGCCGCTTGACCGGACCAAAATCCATGTTGTAAGC 581
197 ThrArgIleLeu...ArgTyrIleLeuArgProGluLysValValIleAl 212
582 AGCTGGCGGACAGCTGCGCTCT...G 604
212 aValGIuAspAsnLysProGluValAlaAlaLeuArgGluValLeuG 229
605 AAAATGCTGCCACATCGAAGACATGATTCGGCGCGCGGATGCTGCG 654
229 lngAlaAsnAspIleGluIleArgValIleProThrLysTyrProSer 245
655 GGTGGTGGACGACGACATTCATTCATGAG...CCGCT 692
246 GlValAlaLysGluLeuIleGluIleLeuThrGlyMetGluValProse 262
693 CGGCGCGAATAAAACCGTGTGACCATTCATTCAGATGTAATACG. 741
262 rGlyGlnArgSerSerIleGlyValLeuMetGlnAsnValAlaThrA 279
742 .....ATTGGCGCTTGTGGCAACAGCGCTGCAACAGCGCGC 783
279 lApheAlaValLysArgAlaIleMetAspAspGluProLeuIleGluArg 295
784 GTGATTCGCTAGGTGCTCTCAAGTCAACAAACCGCGCTCTTGGCTAC 833
296 ValValThrLeuThrGlyAspLysValArgHisLysGlyAsnTyrTrpVa 312
834 CGTTTGGGTGCGAAAGTATCGAAATTCATTCGGCGCGGATTCGTTGACA 883
312 lArgLeuLysThrProIleTyrGlnLeuLeuGln..... 323
884 CAGACACCGCGGTATTCGCGTATTCGAGCGCGGATTCACAA 933
323 ..... 323
934 GCGCGCGCATTTATTTGGACGCTACACATACATTCCTCCATTACGA 983
324 ...GlnValAspTyr.....HisTyrAspAspArgPheProValIlePhe 337
984 AGAAGCGCGGACGAAGAGCTGTTCGGTGGTGGCGCGCGGACGCGACA 1033
337 tGlyLeu.....PrometMetGlyPheIleLeuProAspLeu...G 350
1034 AATATCTCATACCGGCTACACCTCGGCAATTCCTGAAAACAAATC 1083
350 lAlaIleProValThrLysMetThr.....AsnLysLeu 360
1084 TTC.....AAGTCAACACAGCGCTCAACGCGCGGACGCGCAT 1124
361 LeuAlaProAspHisPheGluTyrAlaProProAlaProGluGlnSerCy 377
1125 GGTGCGATGCTGCTACCTAGACGCGGCTGATGCTTGGATATCTGCCCCA 1174
377 sIleArgCysSerAlaCysSerAspAlaCysProValSerLeuMetProG 394
1175 CCCTGCTTTTGGCGCATTAATCGTCGGGATACGAGCGGAGGCA 1224
394 lngLysLeuTyrTrpPheAlaArgSerGluAspHisGluLysSerGlu 410
1225 TTGGGT.....TGCTGAATTCGACGAAGAAGACCTCGCTTTGTG 1265
411 TyrAlaLeuLysAspCysIleGlu.....CysGlyLeuCy 422

```

```

1266 CAGCTTCGCTGCCGCGGCAATACGAATACGCGCGCTGTGCGC.... 1311
422 sAlaTyrAlaCysPro.....SerHisIleProLeuIleGlnTyrP 436
1312 .....AAGTCTCGAAGCAACCATTCGAGAG 1335
436 heArgGlnGluLysAlaLysIleTrpGluIleGlnGlnLys 449
seq_name: Swissprot_40:N121_HUMAN
ID document_block:
ID N121_HUMAN STANDARD: PRT; 1229 AA.
AC Q9Y2N3; 075115; 09Y4S7;
DE 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
DE NUP121 OR KIA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
PI [1]
RN [1]
RP CORDON M., BAUER C., HOLMES A.;
RA Cordes M., Bauer C., Holmes A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 243-1229 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403680; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176(1998).
[3]
RN SEQUENCE OF 1130-1229 FROM N.A.
RP TISSUE=uterus;
RC Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE (BY SIMILARITY).
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NUP1 AND MAMMALIAN P62 AND NUP153.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
-----
DR EMBL; AC006014; AAD28064.1; -
DR EMBL; AB014518; BAA31593.1; -
DR EMBL; AL080109; CAB45713.1; -
KM Nuclear protein; Transpore; Transmembrane; Repeat.
FT DOMAIN 1 40 CISTERNA SIDE (POTENTIAL).
FT TRANSMEM 41 61 POTENTIAL.
FT DOMAIN 62 1229 PORE SIDE (POTENTIAL).
FT DOMAIN 4 10 POLY-ALA.
FT DOMAIN 51 56 POLY-ALA.
FT DOMAIN 294 299 POLY-LYS.
FT DOMAIN 441 444 POLY-SER.
FT DOMAIN 499 502 POLY-PRO.

```

[illegible]

Alignment Scores:	
Quality	102.50
Ratio	0.539
Percent Similarity	41.575
<hr/>	
Length: 457	
Gaps:	18
Percent Identity	21.225

alignment_block:

05-03-2016-125/Rev x N121_HUMAN ..
Align seg 1/1 to: N121_HUMAN from: 1 to: 1229

[illegible]

```

708 1ePheThrAlaProProIysSerGluYsGluGlyProThrProProGly 722
      ||||| :|||
709 1ePheThrAlaProProIysSerGluYsGluGlyProThrProProGly 722
      ||||| :|||
905 CCGGAATACACCGCGGTCTGTGTGATCAACCAATTCGCCCGAGAAATTTG 856
      ||||| :|||
725 ProSerValThrAlaThrAlaProSerSerSerSerLeuProThrThr 744
      ||||| :|||
855 CGATATCTTCACACCAAAACGGTACCGAAGAGCGCGTTGTGTGACTT 806
      ||||| :|||
741 rSerThrThrAlaPro..... 748
805 GAGAACCAACCTTAGGCAATCACGCGCTCGGTGTACAGCGCCGTGCA 756
      .....
746 ..... 746
755 AACAAAGGCCCATGTGTAATTCATTGATTAATGTAGTGCCACACGCT 706
      .....
746 ..... 746
705 TTTATTCGGCGCGACCGCGCTCATGAAATCAATATGCGTCCACTCAAC 656
      ||| ||||| :|||
747 .ThrPheGlnProValPheSer..... 753
655 CCGGACGAGTACGGGCGCCCAATTCATATGTTCGATGTTCGACGAGATT 606
      :: ||||| || || || :||| |||
754 .....SerMetGlyProProIaSer..ValProLeuProAlaProPhe 767
605 TCAGACGCGCACGTCGTGGCGCAGCTGCCTTACAACAATGATTTTGGCTTC 556
      ||||| :|||
768 PheIysGlnThrThrThrProAlaThrAlaProThr..... 779
555 GGTCAACGGCTCAATATACCAACAGCGCGCGTTGAAATCCTCGCGCGCT 506
      .....ThrThrAlaP 783
780 ..... 780
505 CTTTGATTAATGACCTGAGGTCGGCAGCCAGCGGATGTGTCATCGCA 456
      ||||| :|||
783 roleuPheThrGlyLeuAlaSerAlaThrSerAlaValAlaProIleThr 799
455 TTGACAGCAATGGCGACGGCTCGGCATCG.....ACGGCAGGAAATTT 412
      :||| :|||
800 SerAlaSerProSerThrAspSerAlaSerLysProAlaPheGlyPheG 816
      :||| :|||
411 GCGTAACGGACGGGTCCGACGGCGCATGCACAAACCGGATTCGATCAGT 362
      :||| :|||
816 YlleSnsSer..ValSerSerSerValSerThrThrThrSerThra 832
361 TCCGCGCGCACTTCCTCCGCGCTAGTTGGCAGC.....GCTTCAGGT 318
      :||| :|||
832 1aThrAlaAlaSerGlnProPheLeuPheGlyAlaProGlnAlaSerAla 848
317 GCGTACGCTCAACATCGATTTTCGTCTTG.....CTTTC 283
      ||| :|||
849 AlaSerPheThrProAlaMetGlySerLlePheGlnPheGlyLysProPr 865
282 AACGGCATTCACGACTGACTAGATAGCGCTTTTCGCCAGGGTAATCG 233
      :||| :|||
865 oAlaLeuProThrThrThrThrValThrThrPheSerGlnSer..... 879
232 CCGGATTTTCCTGTAACCGCGCGCGAGTAACACACGAG..... 195
      ||||| ||| :|||
880 .....LeuProThrAlaValProThrAlaThrSerSerSerAlaAla 893
194 .....CCCGGATTCCTTTGCTCTTCAACACGACATCGCCTTTTGTGAC 151
      ||||| :|||
894 AspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlaThrSe 910
150 GGCATTCGCTTCCTTGACTTTCATTCAGAGGGCGCATACCGGCAATTCTT 101
510 rSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPheAsnI 927
100 CGCAACCAACGCGAATCTGGTATAGCGCGCGCTGCTAAACGCGCTTCG 51
      ||| :|||

```


927 leprohegylserSerAlaLysSerProLeuProSerTy-ProGlyAla 943

50 TCCGCTCCGCGCGATGGGC 30
::: |||||::: |||

944 AsnProGlnProAlaPheGly 950

seq_name: SwissProt_40:MUra_ECOLI

seq_documentation_block:

ID MUra_ECOLI STANDARD; PRN; 419 AA.
AC P28909;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)
DE (Enoylpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl
transferase) (EPT)
GN MUra OR MUra OR B3189.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-10.
RC STRAIN=K12 / AB1157;
RX MEDLINE=92380955; PubMed=1512209;
RA Margardt J.L., Siegele D.A., Kolter R., Walsh C.T.;
RT Cloning and sequencing of Escherichia coli murZ and purification of
its product, a UDP-N-acetylglucosamine enolpyruvyl transferase.";
RL J. Bacteriol. 174:5748-5752(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1233-1238(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O26 / NGY47;
RX MEDLINE=99216887; PubMed=10103182;
RA Horii T., Kimura T., Sato K., Shibayama K., Ohta M.;
RT "Emergence of fosfomycin-resistant isolates of Shiga-like
toxin-producing Escherichia coli O26.";
RL Antimicrob. Agents Chemother. 43:789-793(1999).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=97148340; PubMed=8994972;
RA Skarzynski T., Mistry A., Monacott A., Hutchinson S.E., Kelly V.A.,
RA Duncan K.;
RT "Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an
enzyme essential for the synthesis of bacterial peptidoglycan,
complexed with substrate UDP-N-acetylglucosamine and the drug
fosfomycin.";
RL Structure 4:1465-1474(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=96151140; PubMed=9485407;
RA Skarzynski T., Kim D.H., Lees W.J., Walsh C.T., Duncan K.;
RT "Stereochemical course of enzymatic enolpyruvyl transfer and
catalytic conformation of the active site revealed by the crystal
structure of the fluorinated analogue of the reaction tetrahedral
intermediate bound to the active site of the C15A mutant of MurA.";
RL Biochemistry 37:2572-2577(1998).
CC -1- FUNCTION: CELL WALL FORMATION. ADDS ENOLPYRUVYL TO UDP-N-
ACETYLGLUCOSAMINE. TARGET FOR THE ANTIBIOTIC PHOSPHOMYCIN.
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-
glucosamine = phosphate + UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-
glucosamine.
CC -1- PATHWAY: FIRST COMMITTED STEP IN PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY. MUra SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M92358; AAA24187.1; -
DR EMBL: U18997; AAA57990.1; -
DR EMBL: AE000399; AAC76221.1; -
DR EMBL: AB028039; BAA78107.1; -
DR PIR: A44917; A44917.
DR PDB: 1UAE; 04-SEP-97.
DR PDB: 1A2N; 29-APR-98.
DR Ecogene: EG11358; murA.
DR InterPro: IPR001986; EPSP_synthase.
DR Pfam: PF00275; EPSP_synthase; 1.
DR ProDom: PD001867; EPSP_synthase; 1.
KW Peptidoglycan synthesis; Cell wall; Cell division; Transferase;
KW 3D-structure; Complete proteome.
FT ACT_SITE 115 115
FT BINDS PERP 6575842255E53F7 CRC64;
SQ SEQUENCE 419 AA; 44817 MW; 6675842255E53F7 CRC64;

alignment_scores: Quality: 102.00 Length: 350
Ratio: 0.662 Gaps: 15
Percent Similarity: 44.000 Percent Identity: 20.286

alignment_block: US-09-303-518D-125 x MUra_ECOLI

Align seg 1/1 to: MUra_ECOLI from: 1 to: 419

67 GCGCCGCCATTAC.....GAAGTCGCTGTCGGCA 101
||||| ::::: |||||
98 GlyProLeuValAlaArgPheGlyGlnGlnValSerLeuProGly 114
102 AGAATATGCGCGTATGCGCC..... 123
||| |||||
114 YcysThrIleGlyAlaArgProValAspLeuHisIleSerGlyLeuGln 131
124TCGATGAAGTCAAGAGGC..... 144
::: |||||
131 InLeuGlyAlaThrIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 147
145 GATGCCGCAAAAGAGCAAGTCGCTGTCGAAGCAAAAGATCCGGC 194
|||::: |||||
148 AspGlyArgLeuLysGlyAlaHisIleValMetAspLysValSerValG 164
195 CGTGGTGTCTACTCGCGCGCTTCAGGCAAAATCGCCGCGATTCACCG 244
||::: ||| ::::: |||||
164 YAlaThrValThrIleMetCysAlaAlaThrLeuAla..... 176
245 GCGAAAGCCGCTACTTCAGTCACTGATTCGCTGTCGAAGCAAGAC 294
|||::: |||||
177GlnGlyThrThr 180
295 GAATCGAGTTGAACGCTACGACCTGAAGCGCTG..... 330
||||| :::
181 IleIleGluAsnAlaAlaArgGlnProGlnLeuValAspThrAlaAsnPh 197
331GCAACTTAAGCGCGAAGAGAGTCCGCCGCAAC 364
||::: |||||
197 eLeuIleThrLeuGlyAlaLysIleSerGlyGlnGlyThrAspArgTleV 214
365 TATATCAA.....TCGGTGTGAGATGCGCTGCGC 396
::: |||||
214 allIleGlyValGlnValArgLeuGlnGlyGlyValTyrArgValLeuPro 230

```

397 ACCGCTCGCTTCAGCAAAATTCCTGCGTATGCGAGCGCTTCCGAT 446
    |||
231 Asparg.....IleuIthrIglYthrPheLeuVal 240
447 CTGCTCATCGCATGACACCAATCCGTGGCTGCGACCTACAGGTCA 496
    |||
240 LAlaVal.....AlaIleSerArgGlyLysIleIle 250
497 TTATCAAGAAGACCGCGAGATTTCAAGCGCGCTGGTGGTATTGAGC 546
    |||
250 IecyAsArgAsnAlaGlnProAspThrLeuAspAla.....ValLeuAla 264
547 CGTTGACCCGACCAAAATCCATGTTTGTAGCAGCTGGCGCAGACG 596
    |||
265 LysLeuArgAsp.....AlaGlyLysAlaSplI 273
597 G.....CCGCTGAAA 607
273 egluValGlyGluAspTrpIleSerLeuAspMetHisGlyLysArgProL 290
608 ATGCTGCCAATCGAAACACATGATTCGCGCGCGCTGCTGCGCGGT 657
    |||
290 YsaIaValAsnValArgThr.....AlaProHisProAlaPhe 302
658 TTGAGTGGCAGCAGCATTCATTTC.....ATCGAGCC 689
303 ProThrAspMetGlnAlaGlnPheThrLeuAsnLeuValAlaGluG 319
690 GGTGCGC...GCGAATAAACCGTGGGAGCATCAATATATCAAGATGAA 736
    |||
319 YThrGlyPheIleThrGluThrValPheGluAsnArgPheMetHisValP 336
737 TTACATTTGCGCGTGTGTTGTCACAGAGCGCTGTACACAGCAGCG 786
    |||
336 rosluLeuSerArgMetGlyAlaHisAlaGlnIleGluSerAsnThrVal 352
787 ATTGCCCTAGTGGTGTCTCAATCAACAACCGCGCTTGGCTGACCGT 836
    |||
353 IleGlyHisGly..... 356
837 TTTGGGTGGGAAAGTATCGCAATTAATTCGCGGCGAATGGTGCACAG 886
    |||
357 .....ValGluLysLeuSerGlyAlaGlnValMetAlaThrA 369
887 ACAACCGCGTGAATTCGCGTTCGCGTATGACGCGCGCATACACAAGC 936
    |||
369 spleuArgAlaSerAlaSerLeuValLeuAlaGlyCysIleAlaGluGly 385
seq_name: SwissProt_40: ODP2_ZYMO

seq_documentation_block:
ID ODP2_ZYMO STANDARD: PRT: 440 AA.
AC 066119:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dihydroliipoamide acetyltransferase component of pyruvate dehydrogenase
DE complex (EC 2.3.1.12) (E2).
GN PDHC OR PDHB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29191 / ZM6.
RX MEDLINE=98175679; PubMed=9515924;
RA Neveling U., Klaesen R., Bringer-Meyer S., Sahm H.;
RT "Purification of the pyruvate dehydrogenase multienzyme complex of
RT Zymomonas mobilis and identification and sequence analysis of the
RT corresponding genes.";
RL J. Bacteriol. 180:1540-1548(1998).

```

```

CC -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydroliipoamide = CoA + S-
CC acetylhydrolipoamide.
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
CC COFACTOR (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC SYMMETRY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: X93605; CAA63808.1; -.
CC HSSD: P10515; IFCY.
CC DR InterPro: IPR001078; 2Oxoacid_dh.
CC DR InterPro: IPR000089; Biotin_lipoyl.
CC DR InterPro: IPR003016; Lipoyl.
CC DR Pfam: IPR004167; e3_binding.
CC DR Pfam: PF00198; 2-oxoacid_dh; 1.
CC DR Pfam: PF00364; biotin_lipoyl; 1.
CC DR Pfam: PF02817; e3_binding; 1.
CC DR ProDom: PD001115; 2Oxoacid_dh; 1.
CC DR PROSITE: PS00189; LIPOYL; FALSE-NEG.
CC KW Glycoyls; Transferase; Acyltransferase; Lipoyl.
CC FT BINDING 43 43 LIPOYL (POTENTIAL).
CC FT ACT SITE 412 412 POTENTIAL.
CC SQ SEQUENCE 440 AA; 46778 MW; 173A98739ACA9A40 CRC64;

```

```

alignment_scores:
Quality: 102.00 Length: 415
Ratio: 0.510 Gaps: 15
Percent Similarity: 48.193 Percent Identity: 21.205

```

Alignment block:

US-09-303-518d-125 x ODP2_ZYMO ..

Align seg 1/1 to: ODP2_ZYMO from: 1 to: 440

```

70 CCGGCCATTACCGAAGTCCGCTTCTGCGGAGCAATATGCGGATGCG 119
    |||
12 ProthrMetThrGlu.....GlyThrIle 19
120 CCCCTCGATGAAGTCAAGAAAGCGATGCCGTCACAAAAGCGCATGCG 169
    |||
19 uAlaIleThrLeuValAlaGlnGlyAspAlaValAlaGlnGlyAspIleIle 36
170 TGTTTGAAGACAAAAGATCCGGCGGTG...TTTACTGCGCGCGGT 216
    |||
36 euAlaGluIleGluThrAspLysAlaIleMetGlnPheGluThrValAsp 52
217 TCAGCAAAATCGCGCGAT.....CACCGGCGGAAAGCGCGTACT 260
    |||
53 AlaGlyIleIleAlaLysIleLeuValProGlnGlySerGluAsnIleAl 69
261 TCATCATGCTGATTCGCGTTGAAGCAACGACGAATCGAGTTTGAAC 310
    |||
69 aValGlyGlnValIleAlaValMetAlaGlnIleAlaGlyGlnAspValSerG 86
311 GTCACGACCTGAGCGCTGCGCAACTTAAGCGGAGGAGCAAGATGCGCGC 360
    |||
86 InValAlaIleAlaSerAlaSerSerGlnIleSerGluProSerGluLysAla 102

```

```

361 AACCTGATCCAAATCCGGTTTGTGACTGCGTCCGACCCGTCGGTTCAG 410
      |||
      |||
103 AspValAlaGln.....LysGluThrAlaAspSe 112
411 CAAATTCCTCCGTCGATGCGGACCGTCGCCATCTTCGTCATGCGGA 460
      |||
      |||
112 rgluThrIleSerIleAspIleSerLysAlaIleSerAsnIleAG 129
461 TGGACACCAATCCGTCGCGGACCCGTCGATTCATTATCAAGAACCC 510
      |||
      |||
129 IyYrGlyAsnLys...ThrGluAsnMetThrAlaSerTyrGlnGlu... 143
511 GCCGAGGATTTCAACGCGCGCTGTGATTGACCCGTTGACCGGACG 560
      |||
      |||
144 .....LysAlaGlyArgIleLysAlaSerProLeuAlaLysAr 156
561 CAAATTCGATGTTGTAAGGACGTCGCGACAGACGTCGCTGGAATG 610
      |||
      |||
156 G.....LeuAlaLysLysAsnH 162
611 CTGCCAATCGAACAACATGATTCGCGCGCCGATCTCCGCGTTTG 660
      |||
      |||
179 LysAlaSplIleGluAlaPheValThrGlyAlaAsnGlnAlaSerSers 195
661 AGTGGACGACATTCATTCATC.....GACCGGTCGCGCGCA 701
      |||
      |||
702 TAAACCGTGTGACCATCAATATCAAGATGTAATTCATTCATTCGCGGT 751
      |||
      |||
195 nProSerValSerThrProGluValSerGlyLysIleThrHisAspThr 212
752 TGTTCGACAGAGCGGCTCG...AACCGGAGCGGCTGATTCG 792
      |||
      |||
212 rOHsAsnSerIleLysLeuSerHisMetArgValIleAlaAsnGly 228
793 CTAGGTGTTCTCAATCAACAAACGCGGCTTCGTCGACGCTT 837
      |||
      |||
229 LeuThrGluSerLysGlnAsnIleProHisIleTyrLeuThrValAspVa 245
838 .....TTGG 841
245 IGlMetAspAlaLeuLysLeuArgSerGluLeuAsnGluSerLeu 262
842 GTGCGAAGTATCCGAAATTCGCGGCGCATTTGTT 879
      |||
      |||
262 IAlaValGlnAsnIleLysIleSerValAsnMetLeuIleLysAlaGln 278
880 .....GACACAGA 887
279 AlalaLeuAlaLeuLysAlaThrProAsnValAsnValAlaPheAspLys 295
888 CAACCGCGTG.....ATTCCGTTTCGGTAT 913
      |||
      |||
295 pOHMetLeuGlnPheSerGlnAlaAspIleSerValAlaValSerValG 312
914 TGAACGCGCGCATTCACACAGCGCGACGATTATTGGGACGCTACCAC 963
      |||
      |||
312 IuGIyGIyLeuIleThrProIleLeuLysGlnAlaAspThrLysSerLeu 328
964 AATTCAGATTTCGTTATCGAAGAGAGCGGATTCGCGCTG 1013
      |||
      |||
329 SerAlaLeuSerVal.....GluMetLysGluLeuIleAlaAr 341
1014 GGTTCGCGCG.....CAGCGGACAAATTCATCCATCAGCGGTACAA 1054
      |||
      |||
341 gAlaHArgGluGlyArgLeuGlnProGlnGluTyrGlnGlyLysThrSers 358
1055 CCCTGGCGCATTTCTGAAAAACAACACTTCAAGTTCAACACAGCCGCTC 1104
      |||
      |||
358 erIleSerAsnMetGlyMetPheGlyIleLysGlnPheAsnAlaValIle 374
1105 AACGGGCGGACGCGGCATGTGTCGCGATGTGTACTTACGAGCGC 1149

```

```

seq_name: SwissProt_40:PUR2_CHICK
375 AsnProGlnAlaSerIleLeuAlaIleLysSerGlyGlnArg 389
      |||
      |||
      |||
seq_documentation_block:
ID PUR2_CHICK STANDARD; PRT; 1003 AA.
AC P21872;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Trifunctional purine biosynthetic protein adenosine-3 [includes:
DE phosphoribosylamine--glycine ligase (EC 6.3.4.13) (GARS) (glycinamide
DE ribonucleotide synthetase) (phosphoribosylglycinamide synthetase);
DE phosphoribosylformylglycinamide cyclo-ligase (EC 6.3.3.1) (AIRS)
DE (phosphoribosyl-aminimidazole synthetase) (AIR synthase);
DE phosphoribosyl-glycinamide formyltransferase (AIR synthase) (GAR
DE transformylase) (5'-phosphoribosylglycinamide transformylase)].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91067455; PubMed=2147474;
RA Alm J., Olu H., Williams J., Zalkin H., Dixon J.E.;
RT "de novo purine nucleotide biosynthesis: cloning of human and avian
RT cDNAs encoding the trifunctional glycinamide ribonucleotide
RT synthetase-aminimidazole ribonucleotide synthetase-glycinamide
RT ribonucleotide transformylase by functional complementation in E.
RT coli.";
RL Nucleic Acids Res. 18:6665-6672(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 5-phosphoribosylamine + glycine = ADP +
CC phosphate + 5'-phosphoribosylglycinamide.
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5'-
CC phosphoribosylglycinamide = tetrahydrofolate + 5'-
CC phosphoribosyl-N-formylglycinamide.
CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide =
CC ADP + phosphate + 5'-phosphoribosyl-5'-aminimidazole.
CC -1- PATHWAY: SECOND, THIRD AND FIFTH STEPS IN DE NOVO PURINE
CC BIOSYNTHESIS.
CC -1- ALTERNATIVE PRODUCTS: THE GARS DOMAIN CAN BE SYNTHESIZED AS A
CC SEPARATE PROTEIN BY ALTERNATIVE SPLICING OF THE GENE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION: BELONGS TO THE GARS FAMILY.
CC -1- SIMILARITY: TO OTHER AIRS AND GART FROM BACTERIA AND EUKARYOTES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X54200; CAA38120.1; -.
CC PIR: S12617; ATCHPR.
CC DR HSSP: P15640; IGSO.
CC DR InterPro: IPR007728; AIRS_related.
CC DR InterPro: IPR001115; GARS.
CC DR InterPro: IPR001555; GART.
CC DR InterPro: IPR002376; formyl_transf.
CC DR Pfam: PF00586; AIRS; 1.
CC DR Pfam: PF02769; AIRS_C; 1.
CC DR Pfam: PF00351; formyl_transf; 1.
CC DR Pfam: PF02842; GARS_B; 1.
CC DR Pfam: PF02843; GARS_C; 1.
CC DR Pfam: PF02844; GARS_N; 1.
CC DR PROSITE: PS00184; GARS; 1.
CC PROSITE: PS00373; GART; 1.

```

KW Multifunctional enzyme; Purine biosynthesis; Ligase; Transferase;
 Alternative splicing.
 FT DOMAIN 1 433 GARS.
 FT DOMAIN 434 805 AIRS.
 FT DOMAIN 806 1003 GART.
 FT ACT_SITE 947 947 BY SIMILARITY.
 FT VARSPEC 434 1003 MISSING (IN ISOFORM GARS-ONLY).
 SQ SEQUENCE 1003 AA; 106544 MW; A0C66BA0EBF791DE CRC64;

alignment_scores:

Quality: 102.00 Length: 490
 Ratio: 0.479 Gaps: 25
 Percent Similarity: 43.469 Percent Identity: 21.429

alignment_block:

US-09-303-518D-125 x PUR2_CHICK ..

Align seg 1/1 to: PUR2_CHICK from: 1 to: 1003

```

7 AAAATCAAAAAGCTTAACCTGACCCATCGCGGCGAGACCGGAGCAAC 56
  ::::::::::: ||| ||| |||
241 GlnIleSerLysAspLeuLeuLysIleArgGluThrValLeuGlnIly 257
  ::::::::::: ||| ||| |||
57 CGTTTAGAGCGCGCGCCATTACCGAAGTCGGCTTGGCGAAGAAAT 106
  ::::::::::: ||| ||| |||
257 sTrhLeuAspGlyMetArgLysGlnIlyLeuProLysGlyValLeu 274
  ::::::::::: ||| ||| |||
107 ATCCGGTATAGCGCCCTGATGAAGTGAAGGCGCATCGCTCAAA 156
  ::::::::::: ||| ||| |||
274 yTrIleGlyLeu.....MetLeuThrLysAspGlyProLysValLeu 287
  ::::::::::: ||| ||| |||
157 AAAGGCCAAGTGTGTTGAAGAC..... 180
  ::::::::::: ||| ||| |||
288 GluPheAsnGlyArgPheGlyAspProGlyLysGlnValIleLeuProLe 304
  ::::::::::: ||| ||| |||
181 ....AAAAGAAATCCGGCGCTGTTTAACTGCGCGCGCTCAGGCAAA 226
  ::::::::::: ||| ||| |||
304 uLeuLysSerAspLeuTyrgluValMetGlnIleValIleAsnLysIle 321
  ::::::::::: ||| ||| |||
227 TCGCCCGCATTCACCGT.....GGCGAAGGCGGTACTTCAGTCACTG 270
  ::::::::::: ||| ||| |||
321 euSerSerSerMetProIleTyrgluAspSerAlaIleValIleThrVal 337
  ::::::::::: ||| ||| |||
271 GTGATTGCGCGTGAAGGCAACGAAATTCAGTTTGAACCTTCAGCAC 320
  ::::::::::: ||| ||| |||
338 ValMetAlaSerGluGly.....TyrProGlyThrTyrPr 349
  ::::::::::: ||| ||| |||
321 TGAAGCGCTG.....GCAAACTTAAGC...GGCGAAGAGTGGCGCGCA 361
  ::::::::::: ||| ||| |||
349 oLysGlyLeuGlnIleThrGlyLeuSerLysAlaLysGlnLeuGlyLeu 366
  ::::::::::: ||| ||| |||
362 ACCGTATCCCAATCCGCTTGTGAGCTGCGCTGCGCACCCGTCGTTGAC 411
  ::::::::::: ||| ||| |||
366 luValAlaPheHisIleGly.....ThrIleLeuLys..... 375
  ::::::::::: ||| ||| |||
412 AAAATTCCTGCGCGATGCGGACCGCTGCGCATCTTCGCAATGCGAT 461
  ::::::::::: ||| ||| |||
376 .....AspGlyLysValValIleThrAsnGlyValIleValIle 387
  ::::::::::: ||| ||| |||
462 GGACACCAATCCGCTGCGCTGCGCACCTTACGATATATCAAGAA... 507
  ::::::::::: ||| ||| |||
387 uThrValThrAlaIleLysGluAspLeuMetThrAlaLeuGlnIleAla 404
  ::::::::::: ||| ||| |||
508 .....GCCCGGAGGATTCAAA..... 525
  ::::::::::: ||| ||| |||
404 snLysGlyValAlaIleAlaIleAsnPheLysGlySerIleTyrglyAsp 420
  ::::::::::: ||| ||| |||
526 .....CGCGCGCTGTTGATTCAGCCGT.....TTGACCGCA 557
  ::::::::::: ||| ||| |||
421 IleGlyTyrglyAlaIleAlaPheLeuSerGlnSerArgGlyLeuThrLy 437
  ::::::::::: ||| ||| |||
558 ACGCAAAATCCATGTTTGTAAAGCAGCTGGC..... 588
  ::::::::::: ||| ||| |||

```

```

437 rLysAsnSerGlyValAspIleAlaIleGlyAsnIleLeuValGlnLysI 454
  ::::::::::: ||| ||| ||| |||
589 .....GCAGACGTGCGCTGTGAATAATCTGCACCATCGAAGA 627
  ::::::::::: ||| ||| ||| |||
454 leLysProLeuAlaIleAlaIleThrSerArgSerGlyCysAsnAlaGlu 470
  ::::::::::: ||| ||| ||| |||
628 CATGAATTCGCGCGC.....CCGCATCC 650
  ::::::::::: ||| ||| ||| |||
471 GlyIlePheAlaGlyLeuPheAspLeuLysAlaIleGlyTyrLysAsp 487
  ::::::::::: ||| ||| ||| |||
651 TCGCGGTTTGAAGTCAGC..... 669
  ::::::::::: ||| ||| ||| |||
487 oIleLeuValSerGlyThrAspGlyValGlyThrLysLeuLysIleAlaG 504
  ::::::::::: ||| ||| ||| |||
670 .....CACATTCATTTTCATGAGCGCGCTGCGCGCAATAA 705
  ::::::::::: ||| ||| ||| |||
504 lValLysLysLysHis.....AspThrIleGlyLysAspLeu 516
  ::::::::::: ||| ||| ||| |||
706 ACCGTGACCATCAATTCATCAAGATGATTAATACATTCGCGCGCT 750
  ::::::::::: ||| ||| ||| |||
517 ValAlaMetCysValAsn.....AspIleLeuAlaGlnGlyAlaGlnPr 531
  ::::::::::: ||| ||| ||| |||
751 .....TTGTTTGAACAGGCGCGCTGGAACACCGAG... 780
  ::::::::::: ||| ||| ||| |||
531 oLeuPhePheLeuAspTyrPheAlaCysGlyLysLeuAspValGluVal 548
  ::::::::::: ||| ||| ||| |||
781 .....CGCGTGAATGCG..... 792
  ::::::::::: ||| ||| ||| |||
548 lAsnIleValIleAlaGlyIleAlaGlnAlaCysGlnLysAlaGlyCys 564
  ::::::::::: ||| ||| ||| |||
793 .....CTAGGTGTTCTCAAGTCAACAACCGCGCTCTTG... 828
  ::::::::::: ||| ||| ||| |||
565 AlaLeuLeuGlyGlyGluThrAlaGlnMetProGlyMetTyrProProG 581
  ::::::::::: ||| ||| ||| |||
829 ....CGTACGTTTGGCGGCAAGATTCGCAATTCCTCGCGCGCAT 874
  ::::::::::: ||| ||| ||| |||
581 yGluTyrglyAspLeuAlaGlyPheAlaValAlaValAlaGlnArgGly 598
  ::::::::::: ||| ||| ||| |||
875 TGCTTGACACAGCAACCGCGATTCGCGTTCGATTCGATTCGAGCGCG 924
  ::::::::::: ||| ||| ||| |||
598 eTrLeuProGlnLeuGlnIleArgIleAlaAspGlyAspValValIleGly 614
  ::::::::::: ||| ||| ||| |||
925 ATTACCAAGGCGCGCATTCATTTTGGAGCGTACCAATCATGATTC 974
  ::::::::::: ||| ||| ||| |||
615 AlaSerSerGlyValHisSer..... 621
  ::::::::::: ||| ||| ||| |||
975 CGTTATCGAAGAGCGCGACCAAGAAGCTGTCGCTGGGTTGCGCGCG 1024
  ::::::::::: ||| ||| ||| |||
621 ..... 621
  ::::::::::: ||| ||| ||| |||
1025 AAGCGGACAAATATCTCATCAGCGTACAAACCTCGGCCATTCCTGAAA 1074
  ::::::::::: ||| ||| ||| |||
622 .....AsnGlyTyrSerLeuValArgIleValGln.....Lys 633
  ::::::::::: ||| ||| ||| |||
1075 AACAAATCTC...TTCAGTTCAACACAGCGGTCAAGGC..... 1110
  ::::::::::: ||| ||| ||| |||
634 SerSerLeuAspPheSerSerGlnValGlyValSerGlyAspGlnThrIe 650
  ::::::::::: ||| ||| ||| |||
1111 .GGGAGCGCGCGCATGCTGCGCATTTGTTACTTACGAGCGCGGTATCCCT 1159
  ::::::::::: ||| ||| ||| |||
650 uGlyAspLeuLeuLeuThrProThrLysIleTyrglySerLysThr..... 664
  ::::::::::: ||| ||| ||| |||
1160 TGGATATCTGCGCACCGCTG 1179
  ::::::::::: ||| ||| ||| |||
665 .....LeuLeuProValLeu 669
  ::::::::::: ||| ||| ||| |||

```

seq_name: SwissProt_40:HFCL_MESAU

seq_documentation_block:

ID HFCL_MESAU STANDARD; PRT; 2090 AA.
 AC P51611;
 DT 01-OCT-1996 (Rel. 34, Created)

01-OCT-1996 (Rel. 34, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
 DE (HCF).
 GN HFC1.
 OS Mesocricetus auratus (Golden hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Mesocricetus.
 CC NCBI_TaxID=10036.
 OK [1]
 RN SEQUENCE FROM N.A.
 RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.:
 Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: UPON LYRIC INFECTION OF PERMISSIVE CELLS, THE HSV
 CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
 CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
 CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
 CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
 CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
 CC SIGNAL (BY SIMILARITY).
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
 CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
 CC AT A DEFINED SITE, PPCE/THEET, WITHIN THE HCF REPEAT (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: DA5419; BAA08258.1; -.
 DR HSP: P02751; 1FNA.
 DR InterPro: IPR003961; FN.III.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF01344; Kelch; 4.
 DR SMART: SM00060; FN3; 2.
 DR Nuclear protein; Repeat.
 KW REPEAT 44 89
 FT REPEAT 93 140 KELCH 1.
 FT REPEAT 148 194 KELCH 2.
 FT REPEAT 217 265 KELCH 3.
 FT REPEAT 266 313 KELCH 4.
 FT REPEAT 313 361 KELCH 5.
 FT DOMAIN 1010 1448 8 X 26 AA APPROXIMATE REPEATS.
 FT REPEAT 1010 1035 HCF REPEAT 1.
 FT REPEAT 1072 1097 HCF REPEAT 2.
 FT REPEAT 1101 1126 HCF REPEAT 3.
 FT REPEAT 1157 1182 HCF REPEAT 4 (DEGENERATE).
 FT REPEAT 1295 1320 HCF REPEAT 5.
 FT REPEAT 1323 1348 HCF REPEAT 6.
 FT REPEAT 1358 1383 HCF REPEAT 7 (DEGENERATE).
 FT REPEAT 1423 1448 HCF REPEAT 8.
 FT SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;
 alignment_scores:
 Quality: 102.00 Length: 462
 Ratio: 0.502 Gaps: 23
 Percent Similarity: 43.939 Percent Identity: 21.861
 alignment_block:
 us-09-303-518d-125/rev x HFC1_MESAU ..

Align seg 1/1 to: HFC1_MESAU from: 1 to: 2090
 1112 CCGCCGTTGAGCGCTGTGTTG.....ACTTGAAGAGTTGTGTTT 1072
 366 ProProAlaArgValGlnLeuValAlaGalaAsnThrAsnSerLeuGluVal 382
 1071 CAGGAATGGCGGAGGCTGTGACCGGTGATGAGTATTTG..... 1032
 382 L...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuLeuGlnLeu 398
 1031TCGGCTGCGCGGACACCGACCGGACGAC 1002
 398 InLysTyrAspLeuProAlaThrAlaAlaThrAlaThrSerProThr... 413
 1001 TCTTTGCTGCGGCTTCTTCATACGAATTCGATTGTGTGCTCC 952
 414ProAsnProValProSerValProAla...AsnPr 424
 951 CAATATATGTCGCGCTGTGTATGCGCGG..... 918
 424 oProLysSerProAlaProAlaAlaAlaAlaProAlaValGlnProLeu 441
 917TTCAATACCGAAGCGGAATTCGCGGTGTGTGTC 879
 441 hrcGlnValGlyLeuThrLeuValProGlnAlaAlaAlaAlaProProSer 457
 878 ACCAATTCG.....CCCGCAATATTCGATCTTGCGAC 841
 458 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 474
 840 CAATACGTTAGCGAGGCGCGGT.....TTGTTCAGTTGAGAC 800
 474 LProThrAlaAlaAlaArgAlaGlnGlyValProAlaValLeuLysValThr 491
 799 CACTTACGGAATACGCGCTG.....GTGTTCAGCGCGCTGTT... 759
 491 LProGlnAlaThrThrGlyThrProLeuValThrMetArgProAlaGly 507
 758 ...GCAACACGACGCAATGTGTAATTCATCTGTAATTCAGTGTCCA 712
 508 GlnAlaGlyLysAlaProValThrValThrSer.....LeuProAlaLe 522
 711 CAGCGTTTATTCGCGCGCGCGCGCTCGATGAATGATGCTGCCAC 662
 522 rValArgMetValValProThrGlnSerAlaGlnGlyThrValIleGly 539
 661 TCAACGCGGAGGATGCGG..... 642
 539 erAsnProGlnMetSerGlyMetAlaLeuAlaLeuAlaAlaAlaAla 555
 641CCGCCAATTCATGTGTTGATGTTGCA..... 612
 556 ThrGlnLysIleProProSerSerAlaProThrValLeuSerValProAl 572
 611GCATTTGAGAGCGCACGTCGCGC 587
 572 agLysThrThrThrValLysThrValAlaValAlaThrProGlyThrThrL 589
 586 CAGCTGCTTACAAATGATTTTGGCTTGGTCAACGCGCTCATAC 537
 589 euProAlaThrValLysValAlaSerSerProValMetValSerAsnPro 605
 536 AACAGCGCGCTTGAATCTCGGCGGCTTGTGTAATGACGCTAG 487
 606 AlaThrArgMetLeuLysThrAlaAlaAlaGlnValGlyThrSerVal 622
 486 GTGCGCAGCC.....ACCG 473
 622 rSerAlaAlaAsnThrSerThrArgProIleIleThrValIleLysSerG 639
 472 GATTGCTGCATGCA.....TTGACGAGATGCGGAC 438
 11 438

```
CC -! SIMILARITY: BELONGS TO THE AMPHIPHYSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60422; CAA42953.1; -.
DR HSSP; P29355; ISEM.
DR InterPro; IPRO03005; Amphiphysin.
DR InterPro; IPRO04148; BAR.
DR InterPro; IPRO04452; SH3.
DR Pfam; PF03114; BAR; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01251; AMPHIPHYSIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Cytoskeleton; SH3 domain; Coiled coil.
FT DOMAIN 10 84 COILED COIL.
FT DOMAIN 144 191 COILED COIL (POTENTIAL).
FT DOMAIN 609 682 COILED COIL (POTENTIAL).
FT SH3.
SO SEQUENCE 682 AA; 75205 MW; 61617F94F38EB20 CRC64;

Alignment_scores:
Quality: 101.50 Length: 371
Ratio: 0.573 Gaps: 17
Percent Similarity: 47.709 Percent Identity: 20.216

alignment_block:
US-09-303-S1B-D-125 x AMPH_CHICK ..
Align seg 1/1 to: AMPH_CHICK from: 1 to: 682

31 CCCATCGGGGAGACCGGAGCAA.....GCCGTTTAGCAGCGCCGGC 74
||| :::::::::::::::::::: :::: |||||
286 ProAlaProAlaArgProLysSerProThgInLeuArgLysGlyProPr 302
75 CATTAACCGAAGTCGCCTTGCTGGCGAGAATAATGCCGGATGCGCCGT 124
::: :::: |||
302 oValProProLeuProLysLeu.....ThProT 312
125 CGATTGAAGTCAGAGGAGCGGATGCCGTCAAAAAAGCCCAACTGCTGT 174
||||| :::::::::::::::::::: :::: |||||
312 hTrLyscIuLeuGIngInGuAsnIleIeAsn.....Leuphe 324
175 GAAGACAAAAGAAATCCGGGCGGTGTACTACGCGCGGCTTAGCGCAA 224
::::: ||| :::: ||:::::::::::::::::: :::: |||||
325 AspAspAsnPhyValProGluLeAsnValThrThrProSerGlnAsnGl 341
225 AATCGCGCCGATTACCGCTGGCGAAAGCGGCTACTTCAGTCAGTCGTA 274
::::: ||| :::: ||:::::::::::::::::: :::: |||||
341 uilePro.....GluThrLysLysValGlnSerLeuLeu. 352
275 TTGCCGTTGAGCAACGACGAATTCGATTGGAAGCGCACGCCCACTGA 324
353 .....AspleuAspPheAspProPheLysProGlu 362
325 GGGCGGCAAACTTAACGGCGAGAGAAGTGCACCACAACCTGATCAATC 374
363 ValValSerThrGlyValThrHisserPrometSerGlnThrLeuPro 379
375 CGGTTGTGAGTGCCTGCGC....ACCGGTCGGTTCGCAAAATATC 418
||||| :::: ||| :::: |||
379 paspleuThrThrThrThrserGluLeuValGlnProLasertGerATRA 396
419 CT.....GCCGTGATGCCGAGCCGTTGCCCATCTTGTCAT 456
306 1st...::: ||||| :::: |||
```

```

457 GCGATGACACCAATCGCTGGCGCCGACCCCTACGTCATTATCAAGA 506
413 GUAANVALTHGLUThLeuThr..... 420
507 ACCCGCCGAGATTCACACCGCGCTGTATGACCGTTGACCG 556
421 .GluValGluValProLeuGluGluLeuValGluValGluThrProT 437
557 AACGCAATTCATGTTCTTAAGCA.....GCGCGCGAGACGCGCG 600
437 hAlaAlaValAlaValGluValGluAlaValLeuAlaGluProAspGluPro 453
601 TCGAAATGCTCCACATCGAACACATGATTCGCGCGCGCGCATCC 650
454 ThGluGluAlaAla.....GluSerIleGluAlaGlyAspGlyGluTh 468
651 TCGCGGTTGAGTGCACGACATTCATTCATCGACGCGCGCGCGCGA 700
468 rThrGlyIleAlaGluValGluSerGluValAlaValSerAlaAlaGlyGly 484
701 ATAAACCGTGTGACCATCAATATCATGATTAATACATTGCGCGT 750
485 .....AlaValAlaValGluAspSerValValAlaAla 496
751 TTGTTGCAACAGCGCGCTGTAACACGCGCGGATTCGCCCTAGTGG 800
497 G1ValaGlyGluGluAlaValAlaThrGluGluGluAlaAlaGluG 513
801 TTCTCAAGTCAACAAACCGCGCTCTGGTACCGTTTGGGTGCGAAG 850
513 Y.....AspLysProGln..... 517
851 TATCGAAATTAAGTGGCGGGAATGTTGACACAGACACCGCGTAT 900
518 .....GlyGluGluValAspValAlaSerGlnGlu 528
901 TCCGTTGCGTATGACGCGCGGATTAACACAGCGCGCATTAATT 950
529 LysValSerSerIleProSerValValIleGluProAlaSerAlaSer 545
951 GGA.....CGTACCAATCATGATTCGTTATGAGAAGGCC 991
545 uGlyGluGluGluGluHis.ValIleMetAsnGluSerLysAspAla 561
992 GCAGCA.....AAGAGCTGTCGCGCTGG 1014
562 AlaAlaGluMetGlyThrGluGlyThrAspSerGluThrSerGlnIleG 578
1015 GTTGGCGCGAGCGGACCA.....AATACTCATACAGCGCTAC 1052
578 ySerGluGluAlaThrGluGluGluIleGlnThrProSerGlnAspG 595
1053 AACCTCGGCC 1063
595 InProAlaSer 598
seq_name: SwissProt_40:CSDB_ECOLI
seq_documentation_block:
ID CSDB_ECOLI STANDARD: PRT: 406 AA.
AC P77444;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Selenocysteine lyase (EC 4.4.1.16) (Selenocysteine reductase)
DE (Selenocysteine beta-lyase) (Sci).
DE GN CSDB OR SUFS OR B1680.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "the complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Stempel G., Seki Y., Sivasubraman S., Tagami H., Takeuchi T.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
[3]
RN SEQUENCE OF 1-10, CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY.
RC STRAIN=K12;
RX MEDLINE=99262626; PubMed=10329673;
RA Mihara H., Maeda M., Fujii T., Kurihara T., Hata Y., Esaki N.;
RT "A nifs-like gene, csdb, encodes an Escherichia coli counterpart of
RT mammalian selenocysteine lyase. Gene cloning, purification,
RT characterization and preliminary x-ray crystallographic studies."
RL J. Biol. Chem. 274:14768-14772(1999).
[4]
RN GENE NAME.
RP STRAIN=K12 / MG1655;
RX MEDLINE=99255563; PubMed=10322040;
RA Patzer S.T., Hanke K.;
RT "Sufs is a Nifs-like protein, and Sufs is necessary for stability of
RT the 2Fe-2S FhuF protein in Escherichia coli."
RL J. Bacteriol. 181:3307-3309(1999).
[5]
RN FUNCTION.
RP MEDLINE=20379012; PubMed=10829016;
RX LaCourciere G.M., Mihara H., Kurihara T., Esaki N., Stadtman T.C.;
RT "Escherichia coli Nifs-like proteins provide selenium in the pathway
RT for the biosynthesis of selenophosphate."
RL J. Biol. Chem. 275:23769-23773(2000).
[6]
RN MOTAGENESIS OF CYS-364.
RP MEDLINE=20206733; PubMed=10739946;
RX Mihara H., Kurihara T., Yoshimura T., Esaki N.;
RT "Kinetic and mutational studies of three Nifs homologs from
RT Escherichia coli: mechanistic difference between L-cysteine
RT desulfurase and L-selenocysteine lyase reactions."
RL J. Biochem. 127:559-567(2000).
[7]
RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RP MEDLINE=20150398; PubMed=10684605;
RX Fujii T., Maeda M., Mihara H., Kurihara T., Esaki N., Hata Y.;
RT "Structure of a Nifs homologue: X-ray structure analysis of CsdB, an
RT Escherichia coli counterpart of mammalian selenocysteine lyase."
RL Biochemistry 39:1263-1273(2000)
[8]
RN FUNCTION: CATALYZES THE RELEASE OF SELENIUM, SO(2), AND SULFUR
RN FROM L-SELENOCYSTEINE, L-CYSTEINE SULFINIC ACID, AND L-CYSTEINE,
RN RESPECTIVELY, TO YIELD L-ALANINE; THE REACTIVITY OF THE SUBSTRATES
RN DECREASED IN THIS ORDER. FUNCTION: THE REACTIVITY OF THE SUBSTRATES
RN IN THE PATHWAY FOR THE BIOSYNTHESIS OF SELENOPHOSPHATE.
RN -1- CATALYTIC ACTIVITY: L-selenocysteine + reduced acceptor = hydrogen
RN selenide + L-alanine + acceptor.
RN -1- COFACTOR: PYRIDOXAL PHOSPHATE.
RN -1- SUBUNIT: HOMODIMER.
RN -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
RN AMINOTRANSFERASES, CSD SUBFAMILY.

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE000263; AAC74750.1; -
DR EMBL; D90812; BAI15463.1; -
DR EMBL; D90811; BAI15457.1; -
DR PDB; 1C0N; 17-JUL-00.
DR PDB; 1TF9; 04-JUL-01.
DR Ecogene; EG13962; csdb.
DR InterPro; IPR000192; AminoTransf_class_V.
DR Pfam; PF00266; aminotran_5; 1.
DR PROSITE; PS00595; AA_TRANSFR_CLASS_5; 1.
KW Lyase; Pyridoxal phosphate; 3D-structure; Complete proteome.
FT BINDING 226 226 PYRIDOXAL PHOSPHATE.
FT ACT_SITE 364 364
FT MUTAGEN 364 364
SQ SEQUENCE 406 AA; 44433 MW; 937AC43C3AD9D8E3 CRC64;
C-XA: LOSS OF CYSTEINE DESULFURIZATION.

alignment_scores:
Quality: 101.00 Length: 362
Ratio: 0.591 Gaps: 17
Percent Similarity: 47.238 Percent Identity: 23.204

alignment_block:

US-09-303-518D-125 x CSDB_ECOLI ..

Align seg 1/1 to: CSDB_ECOLI from: 1 to: 406

229 GCCCGGATTCACCGTGC.....GAAA 251
52 AlaAlaValInIstArgGlyLeuIstHrLeuSerAlaGlnAlaHrGlu 68
252 G.....CGCGTACGTGACGTGATTCGCGTGAAGCA 289
68 SmetGlnAsnValArgGlySerGlnPheIleAsnAlaArgSer 85
290 ACGAGCAATGCGATTTGAACGCTACGCACTGAGCGG..... 330
85 IagGlnIleValPheValArgGlyThrThrGlnIleGlnLeuVal 101
331 GCAACCTTAAGCGCGCAAGAAGTGGCGCG.....AAGCTGATCCA 371
102 AlaAsnSerTrpGlnSerAsnValArgAlaGlyAspAsnIleIle 118
372 ATCCGGTTGTGACGTGCGCGCGCGCGCGTGCAGCAAAATTCG 421
118 eSerGlnMetGlnInIstAlaAsnIleValProTrpGlnMetLeuGly 135
422 CC...GTGATGCGCGAGCGCGTGCATGTCGATGCGATGAGCACC 468
135 IaaGlyAlaGlyAlaGlyLeuArgValIleProLeuAsnProAspGlyThr 151
469 AATCCCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 518
152 LeuGlnLeuGlnThrLeuProThrLeuPhe.....As 162
519 TTTCAAACCGCGCGTGTGATGAGCGCGTTCGAGCA..... 558
162 rGlnGlyThrArgLeuAlaIleThrHisValSerAsnValLeuGly 179
559CGCAAAATCCATGTTTGAAGCGAGCTGGCGCA 591
179 hGlnAsnProLeuAlaGlnMetIleThrLeuAlaInIstGlnHisGlyAla 195
592 GACGTGCGCGTGAATAATGCTGCGCAATGCGCAACATGAA 633
196 LysValLeuValAspGlyAlaGlnAlaValMetHisIstProValAspVal 212

634TTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
212 IGlAlaLeuAspCysAspPheTyrValPheSerGlyHisLysLeuGly 229
656 GTTTCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
229 LyrThrGlnGlyLeuGlyIleLeuTyrValLysGlnAlaLeuGlnGly 245
694GCGCGCAATTAACCGTGGAGCGCGCGCGCGCGCGCG 728
246 MetProProTrpGlnGlyGlyGlySerMetIleAlaThrValSerLeu 262
729 AATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 769
262 rGlnGlyThrThrThrThrThrThrThrThrThrThrThrThrThr 279
770 TGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 819
279 rGlnThrGlyGlyIleIleIleIleGlyLeuGlyAla..AlaLeuGlnTyrVal 295
820 CGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 869
295 eAlaLeu.....GlyLeuAsnIleAlaGlyTyr..... 305
870 GCAATGCGTGAACAGACAGACCGCGCGCGCGCGCGCGCGCGCG 919
306GlnGlnAsnLeuMetHisTyrAlaLeuSerGln 316
920 GCGCGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
316 nLeuGlnSerValProAspLeuThrLeuTyrGly.....ProGlnAsn. 330
970 ATTCGCTTAACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 1019
331ArgLeuGlyVal 334
1020 GCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1069
335 ILeAlaPheAsnLeuGlyLysHisIstAlaTyrAsp.ValGlySerPhe 351
1070 TGAATAACAACTCTCAAGTTCACACAGCGCGCGCGCGCGCGCG 1119
351 eAspAsn.....TyrGlyTyrAlaValArgThrHisHis 363
1120 GCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1152
364 CysAlaMetProLeuMetAlaTyrTyrAsnVal 374

THE UNIVERSITY OF CHICAGO

1961